A HYBRID APPROACH FOR FULL FRAME LOSS CONCEALMENT OF MULTIVIEW VIDEO

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

ÇAĞDAŞ BİLEN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

JULY 2007

Approval of the Thesis

"A HYBRID APPROACH FOR FULL FRAME LOSS CONCEALMENT OF MULTIVIEW VIDEO"

Submitted by **ÇAĞDAŞ BILEN** in partial fulfillment of the requirements for the degree of **Master of Science in Electrical and Electronics Engineering** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences** _________________

Prof. Dr. İsmet Erkmen
Head of Department, **Electrical and Electronics Engineering** _________________

Assoc. Prof. Dr. Gözde Bozdağı Akar
Supervisor, **Electrical and Electronics Engineering** _________________

Examining Committee Members:

Assoc. Prof. Dr. Aydın Alatan(*)
Electrical and Electronics Engineering, METU _________________

Assoc. Prof. Dr. Gözde Bozdağı Akar(**)
Electrical and Electronics Engineering, METU _________________

Assoc. Prof. Dr. Temel Engin Tuncer
Electrical and Electronics Engineering, METU _________________

Assist. Prof. Dr. Çağatay Candan
Electrical and Electronics Engineering, METU _________________

Ayhan Yılmaz (M.S.)
Aydın Yazılım ve Elektronik Sanayii A.Ş. _________________

Date: _________________

(*) Head of Examining Committee
(**) Supervisor

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name	:
Signature	:

ABSTRACT

A HYBRID APPROACH FOR FULL FRAME LOSS CONCEALMENT OF MULTIVIEW VIDEO

Bilen, Çağdaş

M.S., Department of Electrical and Electronics Engineering

Supervisor: Assoc. Prof. Dr. Gözde Bozdağı Akar

July 2007, 77 pages

Multiview video is one of the emerging research areas especially among the video coding community. Transmission of multiview video over an error prone network is possible with efficient compression of these videos. But along with the studies for efficiently compressing the multiview video, new error concealment and error protection methods are also necessary to overcome the problems due to erroneous channel conditions in practical applications.

In packet switching networks, packet losses may lead to block losses in a frame or the loss of an entire frame in an encoded video sequence. In recent years several algorithms are proposed to handle the loss of an entire frame efficiently. However methods for full frame losses in stereoscopic or multiview videos are limited in the literature.

In this thesis a stereoscopic approach for full frame loss concealment of multiview video is proposed. In the proposed methods, the redundancy and disparity between the views and motion information between the previously decoded frames are used to estimate the lost frame. Even though multiview video can be composed of more than two views, at most three view are utilized for concealment. The performance of the proposed algorithms are tested against monoscopic methods and the conditions under which the proposed methods are superior are investigated. The proposed algorithms are applied to both stereoscopic and multiview video.

Keywords: H.264, Video Compression, Error Concealment, Full Frame Loss, multiview video, stereoscopic video

ÖZ

ÇOK GÖRÜŞLÜ VİDEODA TÜM ÇERÇEVE KAYBI İÇİN KARMA BİR YAKLAŞIM

Bilen, Çağdaş
Yüksek Lisans, Elektrik ve Elektronik Mühendisliği Bölümü
Tez Yöneticisi: Doç. Dr. Gözde Bozdağı Akar

Temmuz 2007, 77 sayfa

Çok Görüşlü video özellikle video kodlamayla uğraşanlar için ilgi çeken araştırma alanlarından biri haline gelmektedir. Çok görüşlü videoların hataya müsayit bir kanaldan iletilmesi bu videoların verimli bir şekilde kodlanması ile mümkün olur. Çok görüşlü videoları sıkıştırma için yapılan çalışmalarla beraber pratik uygulamalardaki veri iletimi hatalarını aşmak için yeni hata engelleme ve hata gizleme yöntemleri de gereklidir.

Paket anahtarlamalı ağlarda paket kayıpları, kodlanmış videonun çerçeveleri içerisinde blok kayıplarına veya tüm bir çerçeve kaybına neden olabilir. Son yıllarda tüm çerçeve kaybınınn verimli bir şekilde gizlenebilmesi için algoritmalar önerilmiştir. Ancak çift veya çok görüşlü videolar için önerilen yöntemler sınırlı sayıdadır.

Bu tezde çok görüşlü videolarda tüm çerçeve kaybında hata gizlemek için iki görüşlü bir yaklaşım önerilmiştir. Önerilen yöntemlerde görüşler arasındaki artıklık ve aykırılık bilgisiyle beraber önceden çözülmüş çerçevelerin zamansal bilgileri de kayıp çerçevenin kestirimi için kullanılmaktadır. Çok görüşlü video ikiden fazla görüş içerebilmesine rağmen hata gizleme sırasında en çok üç görüş kullanılmaktadır. Önerilen algoritmaların başarımı tek görüşlü videoda kullanılan yöntemlerle karşılaştırılmış ve hangi durumlarda önerilen algoritmaların daha başarılı olduğu incelenmiştir. Önerilen algoritmalar hem iki hem de çok görüşlü videolarda uygulanmuştır.

Anahtar Kelimeler: H.264, Video Sıkıştırma, Hata Gizleme, Tüm Çerçeve Kaybı, çok görüşlü video, çift görüşlü video

ACKNOWLEDGMENTS

I would like to thank to my supervisor Assoc. Prof. Dr. Gözde Bozdağı Akar for her sincerity, guidance, support and encouragement throughout my graduate study. She is the most helpful and . supervisor that a student can hope for.

This thesis work is supported by 3 Dimensional Television (3DTV) Project of 6. Framework of European Union. I would like to acknowledge the project members for their support and cooperation.

I would like to acknowledge The Scientific and Technological Research Council of Turkey (TUBITAK) for their funds.

I also would like to express my appreciation to the members of Multimedia Research Group (MMRG) who provided me help and information on technical subjects. I also appreciate the great time and fun we had together.

My dear friends from the METU Aikido Dojo also deserve my thanks for their support and warm company. Having friends that are there for you when you need it is a great feeling and makes the life much more enjoyable.

Finally I would like to express my deepest thanks to my loving family for their continuous support and trust in me throughout my life.

To my loving mother,

who always believed in me and made me who I am,

and to my beloved sister,

with whom my life became all beautiful and enjoyable...

TABLE OF CONTENTS

ABSTRACT iv
ÖZ vi
ACKNOWLEDGMENTS viii
DEDICATON ix
TABLE OF CONTENTS x
LIST OF FIGURES xiii
LIST OF TABLES xvii
LIST OF ABBREVIATIONS xx
CHAPTER

1 INTRODUCTION 1
1.1 General 1
1.2 Scope of the Thesis 3
1.3 Outline of the Dissertation 4

2 MULTIVIEW VIDEO CODING 5
2.1 General 5
2.2 MMRG Multiview Codec 8
2.3 H.264 Multiview Extension 10

3 ERROR CONCEALMENT ALGORITHMS 13
3.1 Introduction 13
3.2 Block Loss Concealment 14

3.3 Full Frame Loss Concealment 23
3.4 Concealment Algorithms For Stereo Video 27

4 THE PROPOSED LOSS CONCEALMENT ALGORITHMS 29
4.1 The Proposed Algorithms on MMRG Multiview Codec 30
4.1.1 Temporal Projection 30
4.1.2 Hybrid Projection 31
4.1.3 Spatial Projection 32
4.1.4 Description of the Proposed Algorithms 32
4.2 The Proposed Algorithms on H.264 Multiview Extension 35
4.2.1 Projections to Form the Disparity Field in Anchor Pictures with SHBpix 37
4.2.2 Projections to Form the Disparity Field in First Level Non-Anchor B Pictures with SHBpix 37
4.2.3 Projections to Form the Disparity Field in Second and Higher Level Non-Anchor B Pictures with SHBpix 40
4.2.4 SHBblk Algorithm 40

5 EXPERIMENTAL RESULTS 43
5.1 Objective Quality Comparison Procedure 43
5.1.1 Comparisons of Methods on MMRG Multiview Codec . . 44
5.1.2 Comparisons of Methods on H.264 Multiview Extension . 48
5.2 Subjective Quality Comparison Procedure 66
5.2.1 Double Stimulus Continuous Quality Scale Test Methodology 66
5.2.2 Comparisons of Methods on MMRG Multiview Codec with DSCQS Test 67
5.2.3 Comparisons of Methods on H.264 Multiview Extension with DSCQS Test 69

6 CONCLUSIONS AND FUTURE WORK 71

REFERENCES 74

LIST OF FIGURES

FIGURES

Figure 1.1 Structure of a general multiview streaming system 1

Figure 1.2 Illustration of multiview camera setup used by KDDI Corporation 2

Figure 1.3 Examples of some camera setups for stereoscopic video . . . 2

Figure 2.1 General encoder and decoder diagrams of H.264 based multiview codecs 6

Figure 2.2 Illustration of basic coding modes in H.264 7

Figure 2.3 The video coding layer and network abstraction layer organization in H.264 encoder 8

Figure 2.4 Multiplexing/demultiplexing structure of input/output streams in MMRG Multiview Codec 9

Figure 2.5 Reference Modes in MMRG Multiview Codec 9

Figure 2.6 Reference Structure of MMRG Multiview Codec Mode 1 when coding as (i)Multiview (ii)Stereoscopic 10

Figure 2.7 General Reference Structure of H.264 Multiview Extension Codec when coding as (i)Multiview (ii)Stereoscopic 11

Figure 2.8 Simplified Reference Structure of H.264 Multiview Extension Codec when coding as (i)Multiview (ii)Stereoscopic . . 12

Figure 3.1 Illustration of the multiframe recovery principle in [1]. In the spatial domain, the spatially interpolated MB (center) in (b) should maximize the boundary smoothness measure at F_0 and the following frames. In the temporal domain, the lost motion vector is selected such that the motion compensated MB maximizes the boundary smoothness measure at F_0 and F_1. The arrows in (b) and (c) indicate where the boundary smoothness is imposed 16

Figure 3.2 Illustration of multiframe spatial error concealment algorithm in [1] 16

Figure 3.3 Illustration of multiframe temporal error concealment algorithm in [1] 17

Figure 3.4 Illustration of smoothing constraints in [2]: an arrow between two samples means that the difference between these two samples occurs in the smoothness measure. (a) Smoothing constraint imposed only on the boundary. (b) Smoothing constraint imposed on each sample in the direction towards its nearest boundary 18

Figure 3.5 A possible coding structure described in [3] for a video sequence with B-pictures 18

Figure 3.6 The illustration of the used motion vectors for the generation of the ICMVs in [4] 19

Figure 3.7 Different choices of neighboring blocks for the averaging top find the lost mv. The block configuration in (c) is proposed in [5] to reduce computational complexity 20

Figure 3.8 The illustration of the boundary matching algorithm in [5] . . 20

Figure 3.9 Two interpolation methods employed for spatial error concealment proposed in [6]. (a) Directional interpolation along a specific edge direction (e.g., 45 degrees). (b) Bilinear interpolation 22

Figure 3.10 Block diagram of the proposed SEC approach in [6] 22
Figure 3.11 Block diagram of the proposed temporal error concealment in [6] 23
Figure 3.12 The overall diagram of the system proposed in [6] 24
Figure 3.13 The example of motion vector extrapolation and extrapolation MBs in [7] 24
Figure 3.14 Estimation of optical flow with temporal regularization in [8] with 2 previous frames 25
Figure 3.15 Schematic of the proposed algorithm in [8] 25
Figure 3.16 Schematic of the proposed algorithm in [9] 26
Figure 3.17 Schematic of the proposed algorithm in [10] 27

Figure 4.1 Illustration of projection methods used in MMRG Multiview Codec. P1 represents the temporal projection, P2 represents the hybrid projection and P3 represents the spatial (inter - view projection). 33
Figure 4.2 The block diagrams of SPpix and SPblk algorithms 34
Figure 4.3 Illustrations of projections for the concealment of anchor pictures in SHBpix algorithm 37
Figure 4.4 The backward projection of the anchor disparity vectors for the concealment of first level non-anchor B-Pictures in different schemes of H.264 Multiview Extension 38
Figure 4.5 The backward projection of the anchor disparity vectors for the concealment of second level non-anchor B-Pictures in different schemes of H.264 Multiview Extension 39
Figure 4.6 The backward projection of the anchor disparity vectors for the concealment of third level non-anchor B-Pictures in different schemes of H.264 Multiview Extension 39
Figure 4.7 The block diagrams of SHBpix and SHBblk algorithms . . . 41

Figure 5.1 Concealment results of Botanical sequence when encoded with General referencing structure of H.264 Multiview Extension (Stereo Pixel = SHBpix, Stereo Block = SHBblk) 58
Figure 5.2 Concealment results of Botanical sequence when encoded with Simplified referencing structure of H.264 Multiview Extension (Stereo Pixel = SHBpix, Stereo Block = SHBblk) 59
Figure 5.3 Concealment results of Balloons sequence when encoded with General referencing structure of H.264 Multiview Extension (Stereo Pixel = SHBpix, Stereo Block = SHBblk) 60
Figure 5.4 Concealment results of Balloons sequence when encoded with Simplified referencing structure of H.264 Multiview Extension (Stereo Pixel = SHBpix, Stereo Block = SHBblk) 61
Figure 5.5 Concealment results of Train sequence when encoded with General referencing structure of H.264 Multiview Extension (Stereo Pixel = SHBpix, Stereo Block = SHBblk) 62
Figure 5.6 Concealment results of Train sequence when encoded with Simplified referencing structure of H.264 Multiview Extension (Stereo Pixel = SHBpix, Stereo Block = SHBblk) 63
Figure 5.7 Concealment results of Heidelberg sequence when encoded with General referencing structure of H.264 Multiview Extension (Stereo Pixel = SHBpix, Stereo Block = SHBblk) . . 64
Figure 5.8 Concealment results of Heidelberg sequence when encoded with Simplified referencing structure of H.264 Multiview Extension (Stereo Pixel = SHBpix, Stereo Block = SHBblk) . . 65

LIST OF TABLES

TABLES

Table 5.1 The stereoscopic videos used in the experiments. The Breakdancers video is a multiview video, only first two views of this video is used for the tests 44

Table 5.2 The average PSNR values of the concealment results of each method for each frame in the video sequences. The best result for each video is shown in bold. The results are in dBs. 45

Table 5.3 The average percentages of intra coded blocks and spatially coded (inter-view compensated) blocks in the frames of each video when encoded with MMRG Multiview Codec (MMRGMC). The average percentage of intra coded blocks in the frames of each video when encoded with H.264 are also presented. . . . 46

Table 5.4 The average algorithm completion times (T_{tot}) and motion vector projection times (T_{mv}) in milliseconds for the algorithms for each tested video. 47

Table 5.5 The average algorithm completion times (T_{tot}) and motion vector projection times (T_{mv}) in milliseconds for the algorithms for some of the tested videos at 320 x 240 resolution. 48

Table 5.6 The average PSNR values of the concealment results of each method for each first level non-anchor B frame in the video sequences when encoded with general scheme of H.264 Multiview Extension. The best result for each video is shown in bold. The results are in dBs. 50

Table 5.7 The average PSNR values of the concealment results of each method for each first level non-anchor B frame in the video sequences when encoded with simplified scheme of H.264 Multiview Extension. The best result for each video is shown in bold. The results are in dBs. 51

Table 5.8 The average PSNR values of the concealment results of each method for each second level non-anchor B frame in the video sequences when encoded with general scheme of H.264 Multiview Extension. The best result for each video is shown in bold. The results are in dBs. 52

Table 5.9 The average PSNR values of the concealment results of each method for each second level non-anchor B frame in the video sequences when encoded with simplified scheme of H.264 Multiview Extension. The best result for each video is shown in bold. The results are in dBs. 53

Table 5.10 The average PSNR values of the concealment results of each method for each third level non-anchor B frame in the video sequences when encoded with general scheme of H.264 Multiview Extension. The best result for each video is shown in bold. The results are in dBs. 54

Table 5.11 The average PSNR values of the concealment results of each method for each third level non-anchor B frame in the video sequences when encoded with simplified scheme of H.264 Multiview Extension. The best result for each video is shown in bold. The results are in dBs 55

Table 5.12 The average percentages of intra coded blocks and spatially coded (inter-view compensated) blocks in the backward direction (indicated B) in the frames of each video when encoded with general (indicated with G) and simplified (indicated with S) schemes of H.264 Multiview Extension. The average percentage of intra coded blocks in forward direction (indicated F) in the frames of each video when encoded with H.264 are also presented. 56
Table 5.13 The average algorithm completion times (T_{tot}) and motion vector projection times (T_{mv}) in milliseconds for the algorithms in H.264 Multiview Extension General scheme for each tested video. 57
Table 5.14 The mean opinion scores of the algorithms on MMRG Multiview Codec from the DSCQS Test with best quality score of 41 68
Table 5.15 The mean opinion scores of the algorithms on H.264 Multiview Extension from the DSCQS Test with best quality score of 43 69

LIST OF ABBREVIATIONS

3D 3 Dimensional

CAb Concealment Algorithm on Blocks

CAp Concealment Algorithm on Pixels

DCT Discrete Cosine Transform

DSCQS Double Stimulus Continuous Quality Scale

EC Error Concealment

GOP Group of Pictures

HD High Definition

HD-TV High Definition Television

ICMV Interpolated Candidate Motion Vectors

ISO International Organization for Standardization

ITU International Telecommunication Union

JM Joint Model

JMVM Joint Multiview Video Model

MAD Mean Absolute Difference

MB Macroblock

MV Motion Vector

MMRG Multimedia Research Group

MMRGMC Multimedia Research Group Multiview Codec

MOS Mean Opinion Score

VCL Video Coding Layer

NAL Network Abstraction Layer

NALU Network Abstraction Layer Unit

POCS Projection onto Convex Sets

PSNR Peak Signal to Noise Ratio

SAD Sum of Absolute Differences

SEC Spatial Error Concealment

SHBpix Spatial Concealment on Hierarchical B-pictures on pixels

SHBblk Spatial Concealment on Hierarchical B-pictures on blocks

SOM Self Organizing Map

SPpix Spatial Concealment on P-pictures on pixels

SPblk Spatial Concealment on P-pictures on blocks

CHAPTER 1

INTRODUCTION

1.1 General

The advances in video coding in recent years made it possible for many applications over internet such as video conferencing, video streaming, video publishing, etc. to become practical. Also demands for high quality multimedia resulted in new standards such as High-Definition (HD) Video and High-Definition Television (HD-TV) at high resolutions.

After the introduction of HD-TV, possible improvements on high quality video are investigated and 3 dimensional (3D) video became a popular research area. Different representations of 3D video such as mesh and texture, video and depth, stereoscopic and multiview video and efficient compression methods for these representations are investigated and several algorithms are proposed. Multiview video coding attracted many researchers since methods for monoscopic video coding can also be

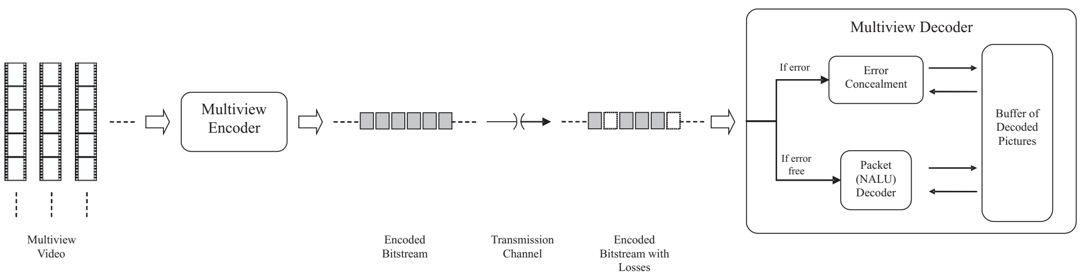


Figure 1.1: Structure of a general multiview streaming system

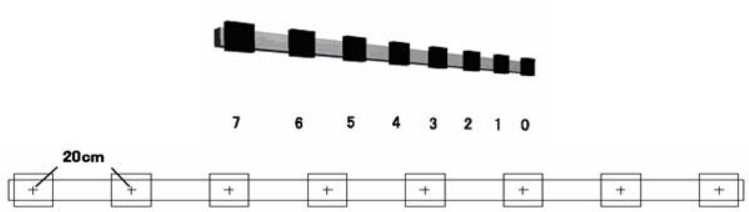


Figure 1.2: Illustration of multiview camera setup used by KDDI Corporation

applied to these videos.

General structure of a multiview streaming system can be seen in Figure 1.1. As it is shown in Figure 1.1, error concealment algorithms are a part of decoder structure which is used in case of errors and loss of packets. In video coding loss of packets may lead to both loss of blocks in a frame or the loss of an entire frame.

Multiview video can be described as multiple video streams shot by several cameras around a single scene. The geometry of the cameras may differ among different setups. 1 dimensional parallel camera setup as in Figure 1.2 is one of the most popular setups. The distance between the cameras is an important parameter which defines the correlation and dependence between the different view points. Consequently this parameter is also important for multiview video coding. Stereoscopic video can be described as a special case of multiview video in which there are only two cameras with a distance of about 3-4 centimeters, the distance between the two eyes of a person. The stereoscopic video simulates the vision of the human visual system when videos from each camera are viewed by the corresponding eye, i.e. video from camera on the left side is viewed by the left eye and vice versa. Some examples to camera

Figure 1.3: Examples of some camera setups for stereoscopic video

setups for stereoscopic video can be seen in Figure 1.3.

Several codec structures based on H.264 video coding standard [11] are proposed for coding of stereoscopic and multiview videos [12, 13]. On November 2006, ITU-T and ISO/IEC JTC1 prepared a draft for the proposal of a new extension to H.264 for multiview coding [14]. However, error concealment methods for those proposed codecs are to be investigated and developed.

As in monoscopic video, the loss of packets during the transmission over a channel may result in the loss of a whole frame in multiview video. Especially in applications regarding low bitrate video, each frame of the video can be sent in a single packet and packet losses lead to frame losses. Also current state of the Joint Multiview Video Model (JMVM) [15], the reference software for the H.264 multiview extension, does not support dividing the frame data into multiple packets, i.e. frames should be sent in a single packet. Consequently loss of a frame should be handled efficiently for proper transmission over error prone channels.

1.2 Scope of the Thesis

The main objective of this study is to develop efficient methods for full frame loss concealment in multiview video. The aim of the concealment algorithms that are proposed is to conceal the lost frames both with high objective quality, i.e. high PSNR, and also high subjective quality, i.e. high perceived quality. The secondary objective is to propose low-complexity algorithms for the full frame loss concealment to be used in practical real-time applications. In order to achieve these objectives, the proposed methods exploit the inter view dependencies in multiview video to conceal the loss of an entire frame.

The performance of the proposed algorithms along with the monoscopic algorithms in the literature are tested on several stereoscopic sequences with different coding methods and objective quality of the methods are compared. Also subjective quality tests are performed in order to measure the perceived quality of the concealment methods.

As being supported by 3DTV project of 6. Frame of European Union, the error concealment methods described in this work will be used along with H.264 Multiview Extension in applications of multiview and stereoscopic video streaming over internet.

1.3 Outline of the Dissertation

The following sections in this thesis are arranged as follows. Chapter 2 describes the two multiview (and stereoscopic) video codecs, MMRG Multiview Video Codec and H.264 Multiview Extension, on which the proposed algorithms are tested.

In Chapter 3, several error concealment methods in the literature for both block losses and full frame losses are described. The methods developed for stereoscopic video are also presented in this chapter.

The proposed multiview concealment methods are described in Chapter 4.

Chapter 5 presents the details of conducted tests along with their results to compare the performances of different concealment methods. The procedure for subjective quality tests are also explained.

Finally, Chapter 6 gives the conclusions to the work done in this thesis and makes suggestions for future work.

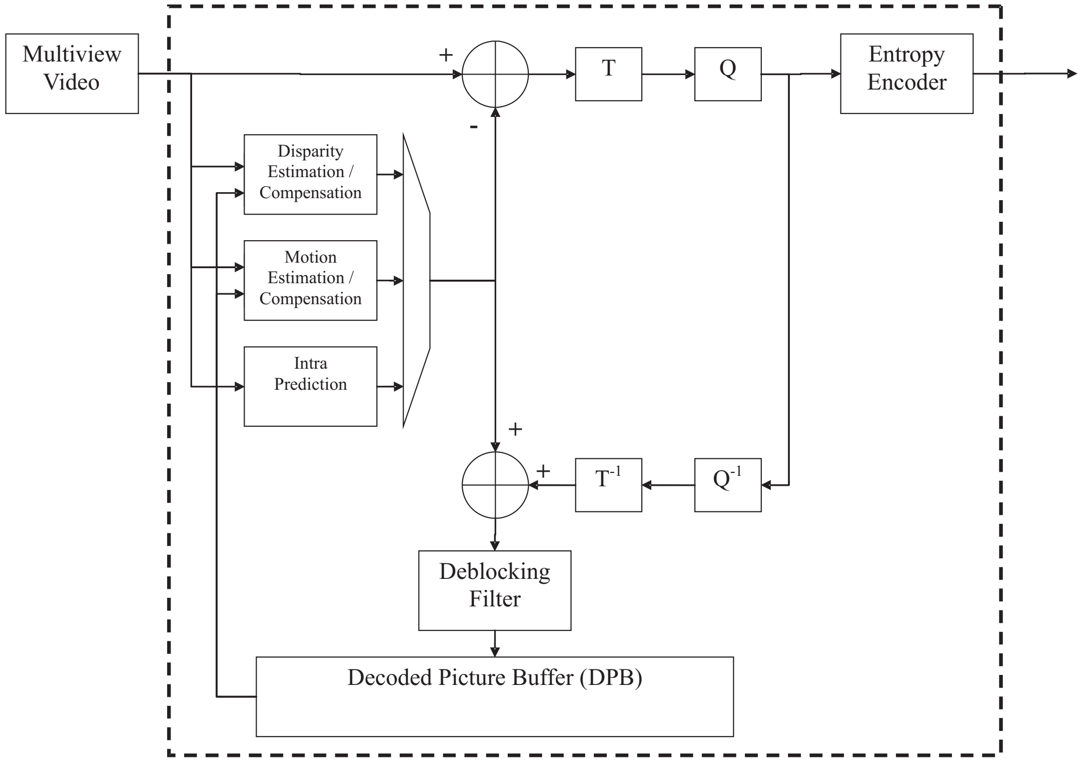
CHAPTER 2

MULTIVIEW VIDEO CODING

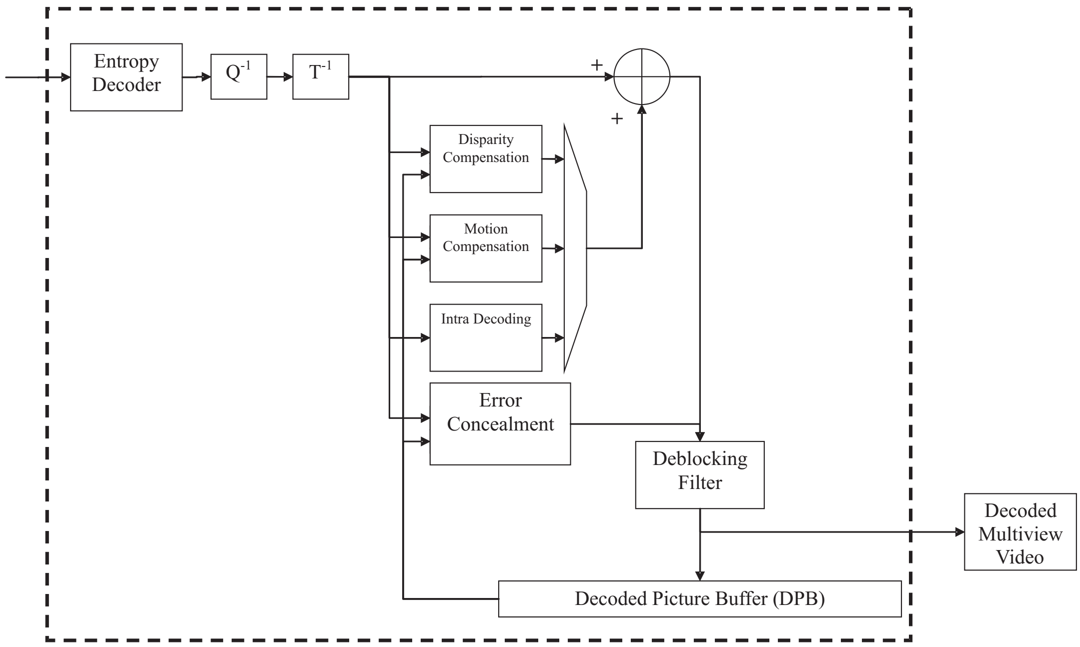
2.1 General

The methods of error concealment are highly dependent on the structure of the codec they are used with since the concealment operation is performed in the decoder in case of errors. The state-of-the art in monoscopic video coding is the H.264 coding standard [11] and most of the methods developed for error concealment are proposed to be used with this codec structure. Most of the proposed codecs for stereoscopic and multiview video also uses the H.264 coding standard as reference. General encoder and decoder diagrams of H.264 based multiview codecs are shown in Figure 2.1. Two of these codecs are used as the base for the proposed concealment algorithms. The MMRG Multiview Codec [13] is an extension to H.264 which exploits inter-view redundancies in addition to temporal redundancies. The H.264 Multiview Extension [14] is the new proposal of ITU-T and ISO/IEC JTC1 for multiview coding and has a much more complex structure compared to MMRG Multiview Codec.

Both of these codecs use H.264 video coding standard as a base and they have almost the same format with the standard H.264 bitstream. In H.264, the frames of the video stream are encoded as intra frames (I-Frames), inter predicted frames (P-Frames) or inter bi-directionally predicted frames (B-Frames). In intra frames each macroblock (MB), 16x16 blocks of pixels in the frame, is encoded as an intra block, i.e. compressed without being predicted from other blocks. Therefore intra frames are self-decodable, i.e. can be decoded without the use of any information from other



(a) General encoder diagram of H.264 based multiview encoders



(b) General decoder diagram of H.264 based multiview decoders

Figure 2.1: General encoder and decoder diagrams of H.264 based multiview codecs

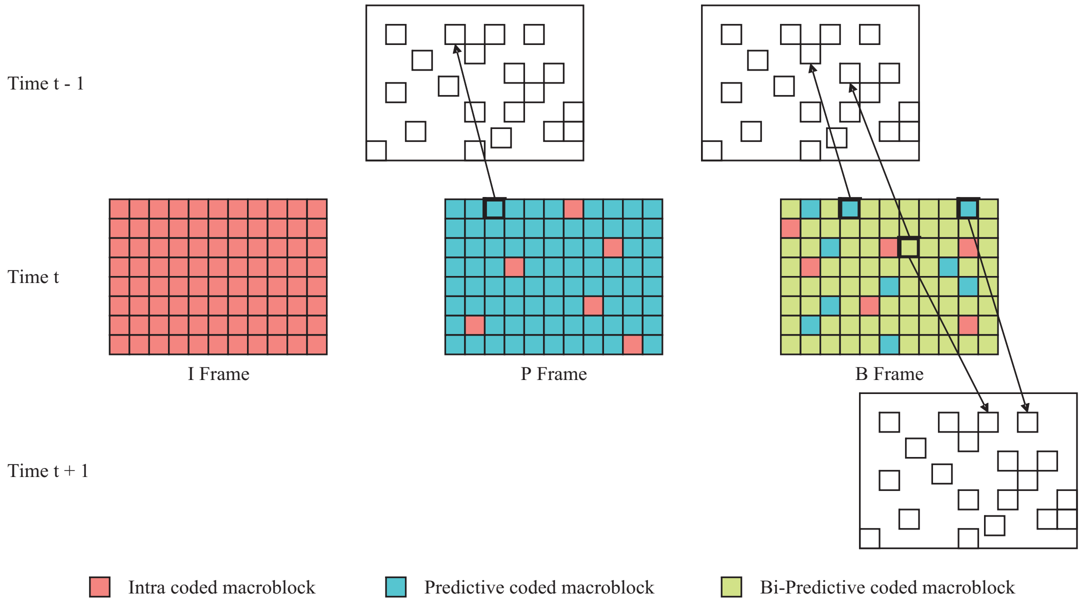


Figure 2.2: Illustration of basic coding modes in H.264

frames. The macroblocks in the P-frames can be encoded as intra blocks or they can also be encoded by being predicted from another pixel block in one of the previously decoded frames. During the encoding of these macroblocks, the closest estimation to the block to be encoded is searched by the encoder in some of the previously encoded frames which are called the reference frames for the frame that is to be encoded. The macroblocks in the B-frames can be encoded as intra, as predicted in P-frames or they can also be encoded by being predicted as a weighted average of two blocks of pixels one of which is from a frame at a previous time instant and the other from a frame at a later time instant. The coding modes of the frames in H.264 is illustrated in Figure 2.2.

The general structure of H.264 is arranged in two layers called video coding layer (VCL) and network abstraction layer (NAL). The transmission related issues and transmission protocol details are handled in the NAL and this layer isolates the video encoding and decoding processes from the transmission process. The general form of the H.264 encoder on VCL and NAL can be seen in Figure 2.3. The encoded bitstream is divided into Network Abstraction Layer Unit (NALU) packets. During the transmission of these packets over a transmission channel, errors may lead to

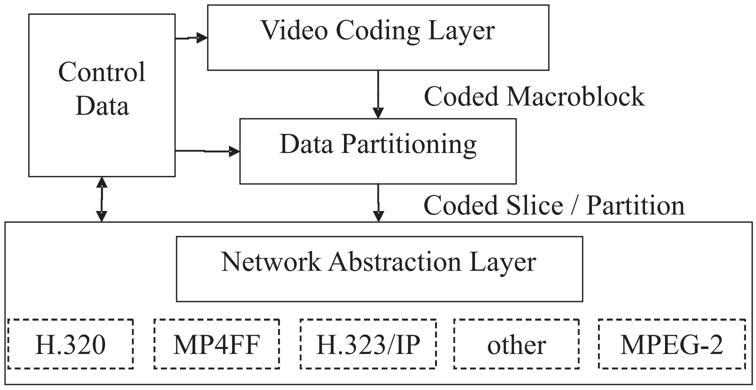


Figure 2.3: The video coding layer and network abstraction layer organization in H.264 encoder

packet losses and these losses are handled in the decoder by error concealment algorithms. Depending on the encoder configuration, the data in the NALU packets can be the encoded data of a group of macroblocks (called slices) or all the macroblocks in an entire frame. Consequently the loss of a NALU packet may lead to loss of a few macroblocks or the loss of whole frame. More detailed information on H.264 standard can be found in [11].

The major difference between the two multiview codecs used in this study is in their referencing structure among the encoded frames which will be explained in detail in the following sections.

2.2 MMRG Multiview Codec

The MMRG Multiview Codec is proposed as an extension to H.264 to encode/decode multiview videos in [13]. The Decoded Picture Buffer (DPB) structure in standard H.264 codec is modified in order to maximize the performance of the codec and the input and output streams of different views are multiplexed/demultiplexed at the input and the output of the decoder as in Figure 2.4.

The MMRG Multiview Codec supports different reference modes to be used in order to provide efficient and flexible coding performance. In Figure 2.5, the referenced frames for the prediction of macroblocks in the P-frame marked in gray in

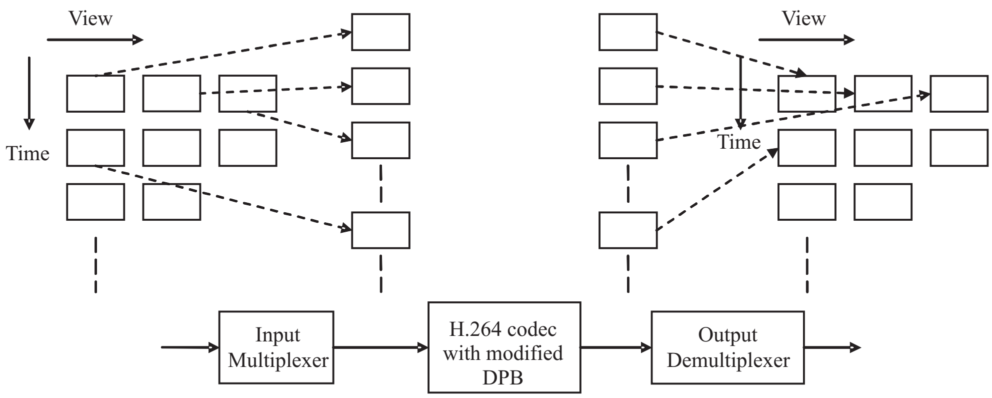


Figure 2.4: Multiplexing/demultiplexing structure of input/output streams in MMRG Multiview Codec

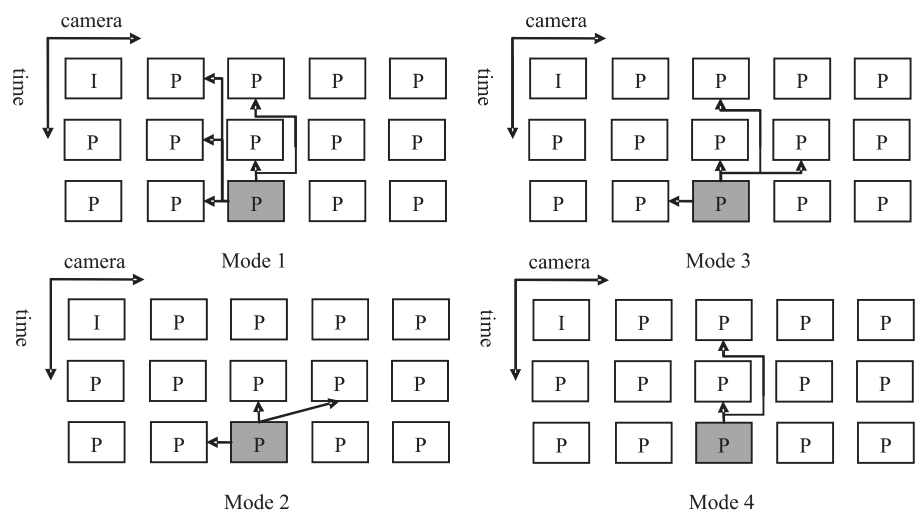


Figure 2.5: Reference Modes in MMRG Multiview Codec

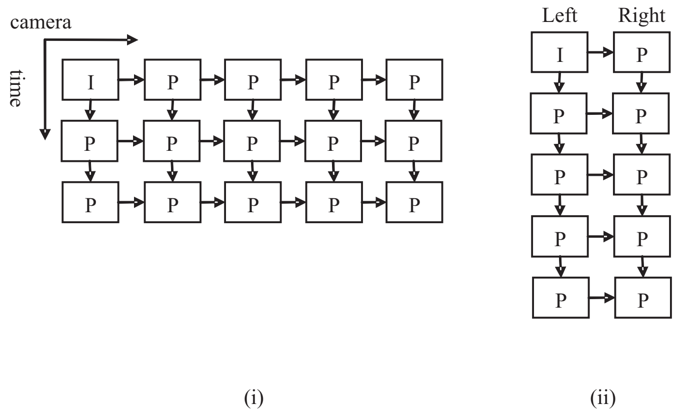


Figure 2.6: Reference Structure of MMRG Multiview Codec Mode 1 when coding as (i)Multiview (ii)Stereoscopic

different modes of MMRG Multiview Codec are illustrated. The mode 1 of this codec is especially important for the scope of this work since it is used for the implementation and test of the proposed full frame loss concealment algorithms. This mode can be used in coding of both stereoscopic and multiview sequences.

The reference structures for coding multiview and stereoscopic videos with mode 1 of MMRG Multiview Codec can be seen in Figure 2.6. It should be noted that although the frames in Figure 2.6 seem to have at most two reference frames, the actual number of reference frames may be more than two as in Figure 2.5.

The MMRG Multiview Codec is chosen as the initial basis for the proposed full frame loss concealment algorithms in this work because of its simpler reference structure compared to H.264 Multiview Extension. Its release is also chronologically earlier than H.264 Multiview Extension.

2.3 H.264 Multiview Extension

First draft of H.264 Multiview Extension is released in November 2006 [14]. The main advantage of this codec is the use of hierarchical B-Pictures in coding of multiview video to remove both temporal and inter-view dependencies. The hierarchical B-Pictures are used to encode the frames in a video stream so that some of the

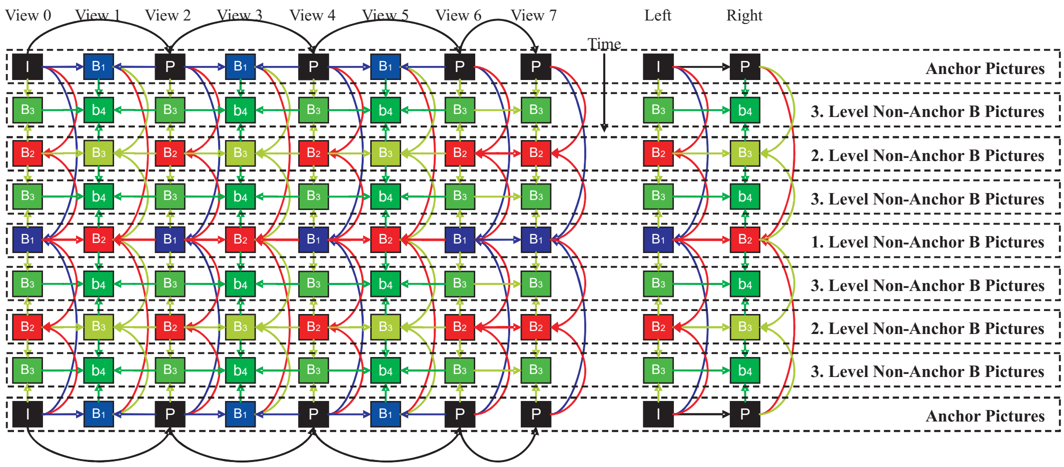


Figure 2.7: General Reference Structure of H.264 Multiview Extension Codec when coding as (i)Multiview (ii)Stereoscopic

frames can be omitted in case of low bandwidth to provide temporal scalability. This idea can be applied to both in temporal and inter-view direction in H.264 Multiview Extension. The general reference structure of a group of pictures (GOP) of H.264 Multiview Extension for multiview and stereoscopic coding is shown in Figure 2.7.

As it can be seen in Figure 2.7, the frames are encoded in different levels. The first frame of each view in the GOP are called anchor pictures and these frames are encoded only with intra mode and inter-view compensation. The non-anchor frames in the GOP can be classified as first, second and third level non-anchor B pictures. The number of levels can be increased with the GOP length.

The use of hierarchical B-pictures provides a good compression efficiency in addition to scalability and error resilience. For instance the **B4** frames (highest level frames) in Figure 2.7 can be removed from the multiview stream without affecting the decoding process of the other frames. Similarly the errors occurring in the **B4** frames does not propagate to the other lower level or same level frames. In general errors in a certain level of hierarchically encoded B-pictures only propagates to the higher levels.

Besides its advantages, the structure of hierarchical B-pictures also have disadvantages. The most important of these is the requirement of extensive memory and

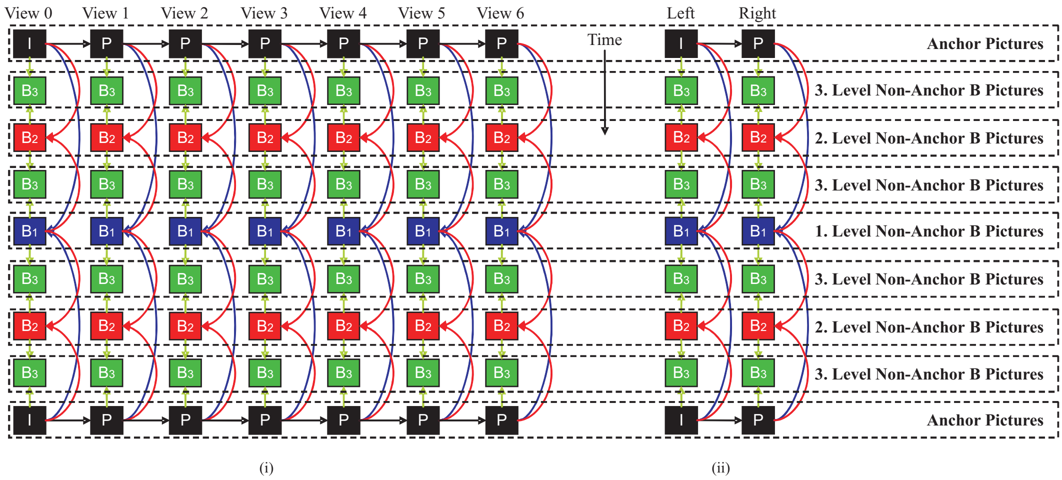


Figure 2.8: Simplified Reference Structure of H.264 Multiview Extension Codec when coding as (i)Multiview (ii)Stereoscopic

high encoder complexity. For example during the encoding and decoding of a GOP, about 64 frames for 8 views, almost all the frames should be kept in memory for the references of other frames in the GOP. Taking this problem into consideration, simpler reference structures are proposed instead of the general reference structure in Figure 2.7. Among several structures, the referencing structure in Figure 2.8 is selected to be the best candidate with the closest performance to general structure providing much more simpler implementations [16, 17]. It can be seen that inter-view references are only at the beginning of the GOPs and the stream is almost simulcast coded.

Although it is based on H.264 standard, current state of the JMVM, reference software of H.264 Multiview Extension, does not support dividing a frame into slices and partitioning into multiple NALU packets yet. Therefore each frame should be sent in a single NALU packet in current implementations. Slices support is planned to be added in later versions.

CHAPTER 3

ERROR CONCEALMENT ALGORITHMS

3.1 Introduction

The transmission of data through a medium, practically through a network, often results in transmission errors. These errors can be handled with a number of different methods such as forward error protection or retransmissions. When an encoded bitstream, specifically video data, is received erroneously in the decoder side, i.e. when error protection or error correction methods were not able to avoid the errors or these methods were not used, the erroneous artifacts in the video are concealed from the user via methods called as error concealment. The error concealment operation is a post processing step in the decoder side and it is the last step to handle the errors.

Error concealment methods are investigated for over 15 years and many different algorithms are proposed in literature for different coding schemes and error types. Although most of the concealment methods on monoscopic video can not be applied to the frame loss case in multiview video which is the main subject of this thesis, the monoscopic loss concealment algorithms in the literature are also reviewed in order to provide background information to the reader on error concealment algorithms and possible exploitations of information in the previously decoded frames.

3.2 Block Loss Concealment

In general, the loss of encoded data of a part of the frame results in loss of some of the macroblocks in the frame. The block losses generally occur as bursts in practice, i.e. consecutive macroblocks are lost at the same time. Some of the major block loss error concealment algorithms are reviewed in this section. The approaches to conceal the lost macroblock information varies on the assumption of the available information to the decoder at the time of the loss. For instance, if no previously decoded frames are assumed to exist, methods to interpolate the available pixels in the frame to be concealed are used which can be generalized as spatial methods. The methods that use the information from the previous frames can be called as the temporal methods. These methods generally apply some boundary matching or block matching techniques to provide continuity in the concealed frame. The motion vectors of the neighboring blocks can also be used if available.

In [18], the error concealment methods developed before 1998 are reviewed and described. The methods reviewed in this paper are divided into 6 categories: motion compensated temporal prediction, spatial and frequency domain interpolation, maximally smooth recovery, projection onto convex sets (POCS) and recovery of coding mode and motion information. In motion compensated temporal prediction the lost macroblocks are replaced by a proper block from the previous frames. In spatial or frequency domain interpolation methods the smoothness property of natural images are used and the lost block information is assumed to be close to the corresponding coefficients in the neighboring blocks. The pixels or DCT coefficients of the lost blocks are estimated by interpolation from these blocks. Maximally smooth recovery methods also use the smoothness property of the videos and replace the missing blocks by a smoothness constraint with the spatially and temporally adjacent blocks. POCS method utilizes the use of convex sets. The convex sets are derived by requiring the recovered block to have a limited bandwidth either isotropically (for a block in a smooth region) or along a particular direction (for a block containing a straight edge). POCS method is an iterative method and uses only spatial information for concealment. Finally in the methods of recovery of coding mode motion information, the motion vectors and coding mode of the lost block is assumed to be similar to the neighboring blocks and estimated.

In [19], an error concealment technique based on an artificial neural network model, model of self organizing maps, is used to estimate the motion vectors of the lost macroblock. The method is proposed to be effective in the estimation of fast and complex motion by utilizing non-linear estimation property of SOMs. The proposed method is found to be effective especially for fast motion videos and it has a low computational complexity.

In [20], a recursive block matching algorithm is proposed. Basically the proposed algorithm conceals the upper and lower halves of the lost macroblock recursively, applying block matching considering the mean absolute differences (MAD). MAD of top neighboring block with blocks from the previous frame and MAD of bottom neighboring block with blocks from the previous frame are taken into account along with the MAD of previous estimate of the half of the lost block with the blocks from the previous frame in the iterations. The method is concluded to be effective fast motion videos and in case of multiple loss of slices in the video frame where other block matching methods are not as successful both in the subjective assessment and PSNR comparisons.

In [21], two algorithms are proposed. The first algorithm is for intra frames for which the temporal information is not available and spatial concealment techniques are used. The proposed algorithm first detects edges components in neighboring boundary pixels, and connects broken edges in the lost macroblock (MB) via linear approximation. Then, the lost MB is partitioned into segments based on the recovered edge information. Finally, each pixel in a segment is directionally interpolated from the boundary pixels that are adjacent to the segment. For the P-frames a MB recovery algorithm with two modes is proposed. The algorithm chooses either to linearly interpolate four pixels from the previous frame or to estimate the lost motion vector by applying block matching with sum of square differences criterion with the left and top neighboring pixels of the lost MB. In rare cases of very fast motion the

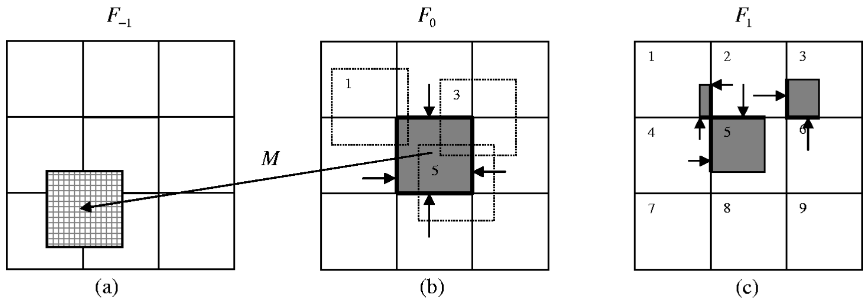


Figure 3.1: Illustration of the multiframe recovery principle in [1]. In the spatial domain, the spatially interpolated MB (center) in (b) should maximize the boundary smoothness measure at F_0 and the following frames. In the temporal domain, the lost motion vector is selected such that the motion compensated MB maximizes the boundary smoothness measure at F_0 and F_1. The arrows in (b) and (c) indicate where the boundary smoothness is imposed

algorithm uses the intra frame concealment technique for the P-frames as well. The P-frame error concealment method uses error tracking and dynamic mode weighting. It conceals a pixel as a weighted sum of candidate pixels that are reconstructed using the above two concealment modes. The weighting coefficients are dynamically determined to reduce the propagation error and the concealment error. It was shown with simulation results that the proposed methods provide significantly better performance in error-prone environments than the conventional concealment methods.

The paper in [1] proposes two multiframe concealment methods to estimate the lost blocks in the video frame while minimizing the error propagation (Figure 3.1).

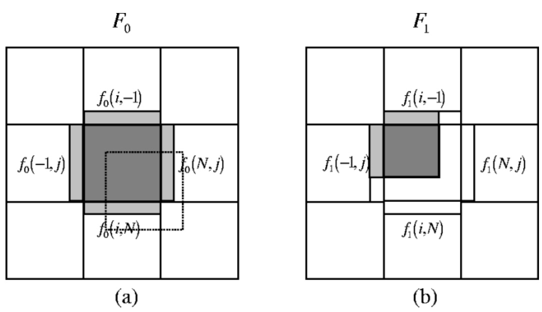


Figure 3.2: Illustration of multiframe spatial error concealment algorithm in [1]

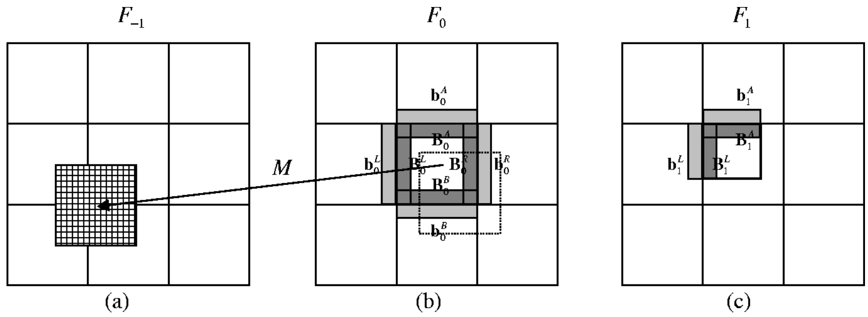
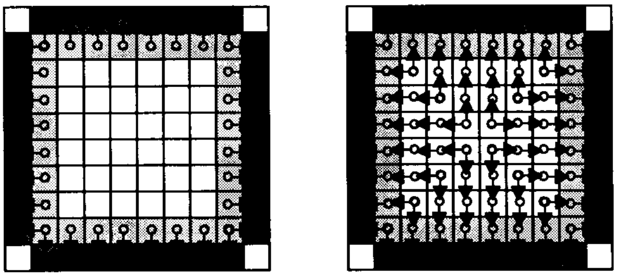


Figure 3.3: Illustration of multiframe temporal error concealment algorithm in [1]

In the spatial method the lost MB is recovered by maximizing a smoothness measure across multiple frames as in the Figure 3.2. In the temporal method, the motion-compensated MB uses an estimated motion vector that exhibits the minimal boundary variation in the current decoded frame as well as in the dependent succeeding inter-coded frames (illustrated in Figure 3.3). The experimental results show that by using multiframe error concealment the quality of the corrupted frames can be improved by 1-4 dB in the tested video sequences, and the error propagation into succeeding motion-compensated frames can also be reduced. The visual improvements are observed to be more significant than the PSNR improvements.

The paper in [2] considers how to reconstruct a damaged block from the received coefficients and its boundary information. The proposed method makes use of the smoothness property of common image signals and produces a maximally smooth image among all the images with the same received coefficients and boundary conditions as in Figure 3.4. It operates by minimizing the inter-sample differences within each damaged block and between adjacent blocks. The boundary information is propagated into the damaged blocks such that the transitions along block boundaries are as smooth as possible. The optimal solution is obtained by two linear transformations, where the transform matrices depend on the loss patterns and can be calculated in advance. When several spatially adjacent blocks are damaged, initial reconstructions of these blocks are first accomplished by the direct inverse transform with the missing coefficients substituted by zeros. Then the proposed technique is repeated several times, using the previously reconstructed values as the boundary information



(a) (b)

Figure 3.4: Illustration of smoothing constraints in [2]: an arrow between two samples means that the difference between these two samples occurs in the smoothness measure. (a) Smoothing constraint imposed only on the boundary. (b) Smoothing constraint imposed on each sample in the direction towards its nearest boundary

in the next step. Simulation results in the paper have shown that the proposed algorithm is very effective for recovering the DC and low-frequency coefficients. When high frequency coefficients are lost, the proposed reconstruction scheme produces results similar to that obtained by replacing the missing coefficients with zeros: both will blur sharp edges.

In [3], another method that relies on motion vector prediction is proposed. The proposed methods are applicable to coding schemes including B-pictures such as in Figure 3.5. For each type of frame in the video (I, P or B), different motion vector projection methods are proposed. Both in P and I-picture motion vector estimation, scaled versions of forward motion vectors of previous B-pictures are proposed to

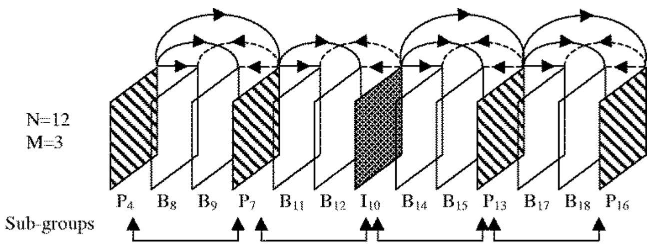


Figure 3.5: A possible coding structure described in [3] for a video sequence with B-pictures

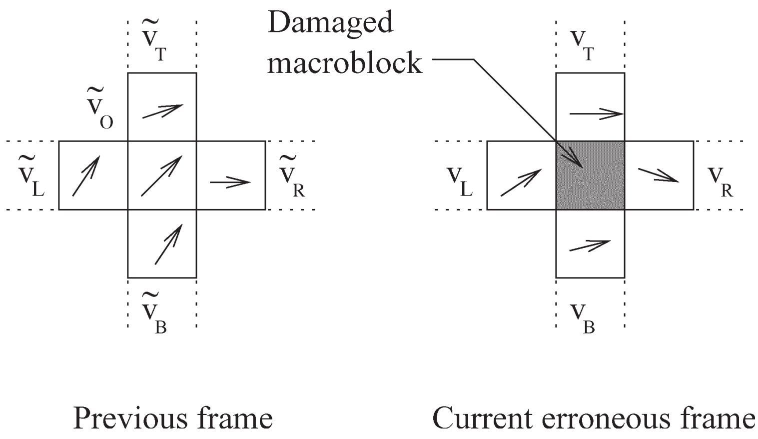


Figure 3.6: The illustration of the used motion vectors for the generation of the ICMVs in [4]

be used. For example for the motion vectors in P7, 3x(motion vectors of B8) or (3/2)x(motion vectors of B9) can be used. As for B-pictures, it is noted that forward, backward or bidirectional projection of the motion vectors can be used. The proposed projection methods can both be used in slice loss or full frame loss cases.

In [4], boundary matching error concealment methods are improved with the introduction of the notion of interpolated candidate motion vectors (ICMV). The candidate motion vectors among which one is chosen as the concealed motion vector with a sum of absolute differences (SAD) criterion are calculated with an interpolation of the motion vectors of the neighboring macroblocks and macroblocks around the corresponding block in the previous frame. The ICMVs are created as follows: the change of magnitude and angle of the top, bottom, left and right neighboring blocks of the lost block (can be seen in Figure 3.6) from previous frame to the current frame is applied to the motion vector of the block corresponding to the lost block in the previous frame and these four motion vectors are taken as ICMVs. A more generalized proposal is to apply such transformation to not only closest neighbors but to a wider area of blocks. Then the boundary matching algorithm may choose among these candidates. The proposed method is concluded to be better than the methods with normal candidate motion vectors especially in high motion videos.

In [22], a set of error concealment schemes based on the latest H.264 video stan-

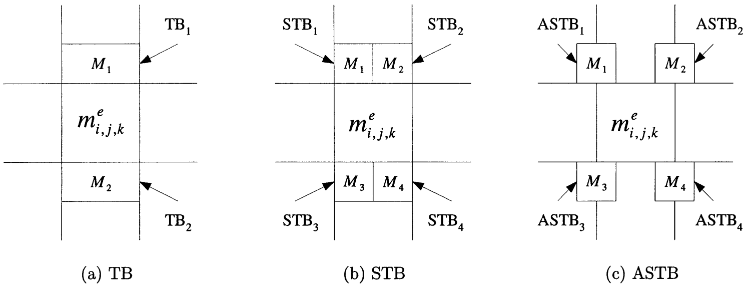


Figure 3.7: Different choices of neighboring blocks for the averaging top find the lost mv. The block configuration in (c) is proposed in [5] to reduce computational complexity

dard to improve error resilience ability for video consumer applications is proposed. The schemes include a method of subblock-based refined motion compensated temporal concealment with weighted boundary match criterion, an algorithm of refined directional weighted spatial interpolation, and an adaptive spatial/temporal estimation method with low complexity to combine the above algorithms. The ability of dealing with detailed motions and object edge integrity is promoted.

In [5], three error concealment techniques are proposed and their performances are investigated. In error concealment using neighboring motion vectors, the motion vector of the lost block is estimated from the four neighboring sub-blocks by averaging their motion vectors as illustrated in Figure 3.7. In the method of extended boundary pixels shown in Figure 3.8, a boundary matching type algorithm is pro-

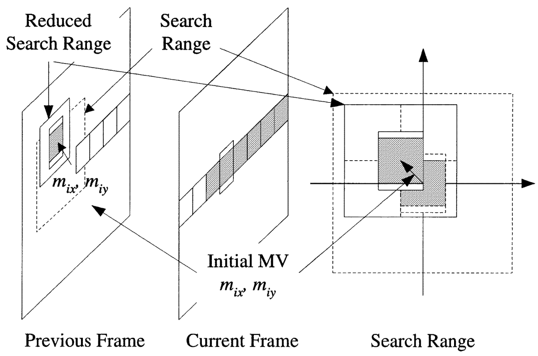


Figure 3.8: The illustration of the boundary matching algorithm in [5]

posed matching the boundaries of two or three sides of the blocks. The use of optical flow fields is proposed as the third method. The optical flow of the neighboring blocks is calculated using the famous Horn and Schunch and the optical flow information is used in concealment. In the comparisons, the boundary matching was the most successful except when there is the loss of multiple lines of video data in which optical flow performed better.

[6] also combines multiple methods and provides a selection criteria among the proposed methods. The spatial error concealment (SEC) module proposed in the paper switches interpolation methods in Figure 3.9 (bilinear or directional) based on the directional entropy in the neighborhood of each missing MB (method diagram shown in Figure 3.10), thus improving upon the performance of both single interpolation methods and previous edge strength based switching approaches. Temporal concealment has been optimized by employing only those features that were found to improve the performance without significantly increasing complexity and the block diagram of the proposed temporal error concealment can be seen in Figure 3.11. The mode selection (MS) algorithm developed examines the suitability of the proposed TEC method for concealing each missing MB, by evaluating the levels of motion compensated activity in the neighborhood of that MB and switching to spatial concealment accordingly. The overall system is illustrated in Figure 3.12. The system is compared to the algorithms in JM decoder and up to 9 dB improvement is observed in the tests.

The paper in [23] presents an adaptive error concealment technique to determine the weighting of spatial interpolation and temporal compensation, and thereby to improve perceptual quality. The proposed temporal prediction for error blocks used median function from the neighboring blocks of the current and the previous frame to achieve high-accuracy motion vector. The proposed adaptive scheme keeps low computational load to make it applicable to real-time MPEG systems. Although simple bilinear method is used for spatial interpolation, the proposed algorithm outperforms the conventional error concealment methods in simulations. It is noted that the performances of individual spatial and temporal concealment techniques can be

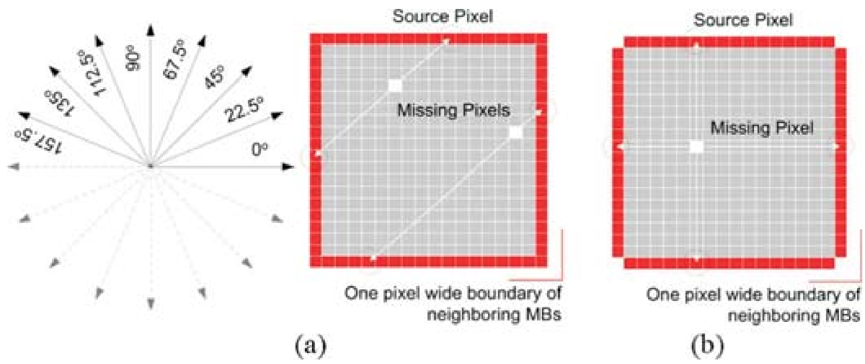


Figure 3.9: Two interpolation methods employed for spatial error concealment proposed in [6]. (a) Directional interpolation along a specific edge direction (e.g., 45 degrees). (b) Bilinear interpolation

improved by using more complex algorithm. The basic system setup is illustrated in Figure 3.13.

Similar to adaptive algorithms, the use of decision trees to choose among multiple error concealment methods is proposed in [24] along with two new temporal error concealment schemes. The first one is based on estimating global pan parameters, and the second one is based on separating a MB into top and bottom halves for separate use of MVs from above and below. These new EC methods were noted to be often the best choices among the fixed methods which can be seen in Figure 3.14. The use of a decision tree to choose adaptively among the various methods consistently provided lower distortion than any fixed method alone according to the results in [24].

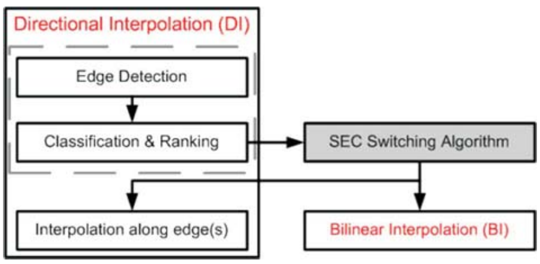


Figure 3.10: Block diagram of the proposed SEC approach in [6]

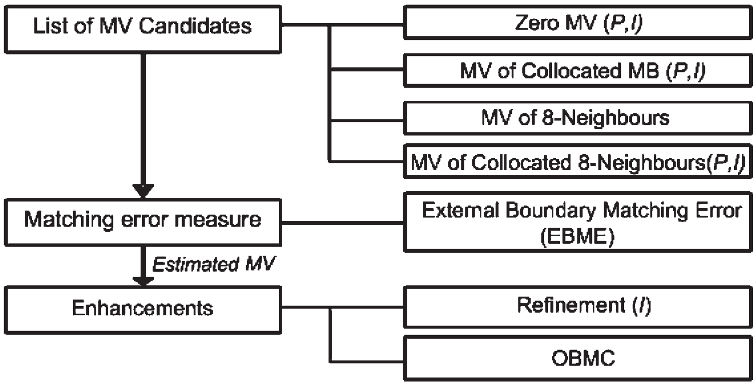


Figure 3.11: Block diagram of the proposed temporal error concealment in [6]

Another algorithm to estimate the motion vector of a lost macroblock is proposed in [7]. The algorithm, which is called motion vector extrapolation, uses a weighted averaging method over some of the motion vectors in the previous frame to assign a motion vector to the lost macroblock. The extrapolation method is illustrated in Figure 3.13. First the motion vectors of the neighboring blocks to the block corresponding to the lost block in the previous frame are projected to the current frame as in Figure 3.13. Second, after the projection, weight for each projected motion vector is calculated as the area of the overlapping region of macroblock compensated by the projected motion vector and the lost block. Finally a motion vector for the lost block is calculated by weighting the projected motion vectors.

3.3 Full Frame Loss Concealment

When entire frame is lost and to be concealed, the only information available to the decoder is the previously decoded frames and their motion vectors. Therefore algorithms such as boundary matching that uses the information of available pixels in the frame to be concealed are not applicable. In the several algorithms proposed in the literature, the motion vectors of the previously decoded frames are used to generate a motion model between the frames and the motion in the frame to be concealed is estimated. As a result a concealed picture is formed by the motion compensation by those motion vectors.

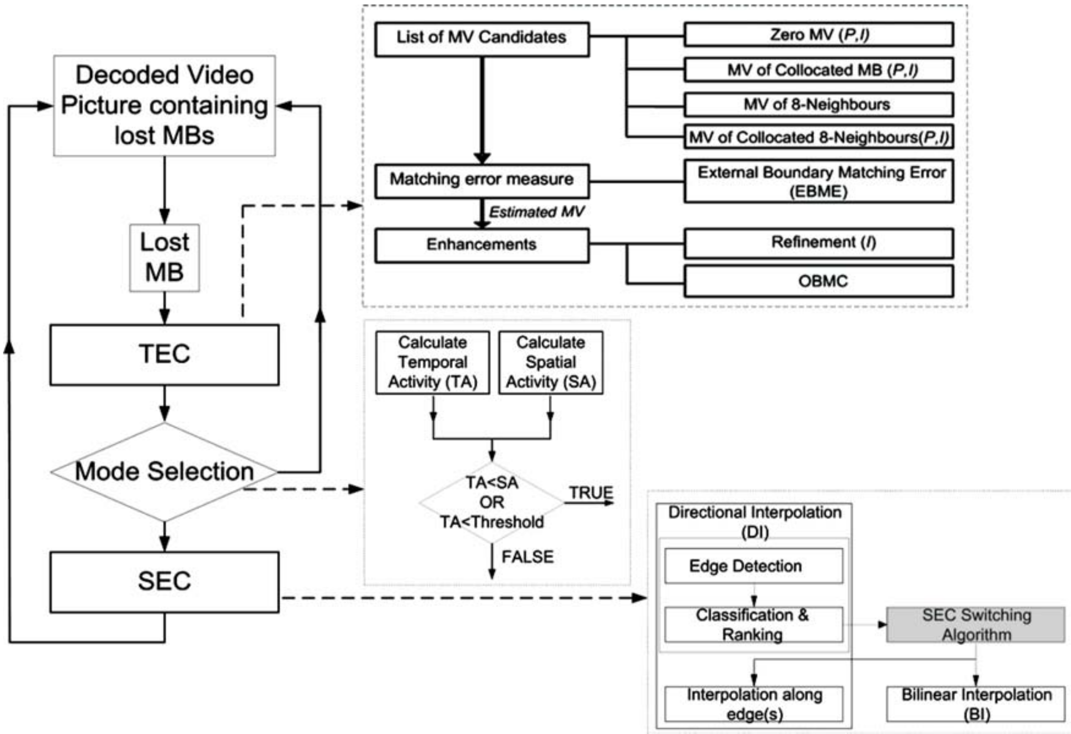


Figure 3.12: The overall diagram of the system proposed in [6]

The methods in [25] are proposed to be used in the H.264 reference software. These two basic concealment methods are frame copy and motion vector copy algorithms. Frame copy algorithm is simply repeating the temporally closest frame to the lost frame. And motion vector copy algorithm is using the motion vector information of the temporally closest frame to the lost frame as the motion vectors of the lost frame and forming the concealed frame with motion compensation. The authors concluded that the motion vector copy algorithm outperforms the frame copy algorithm

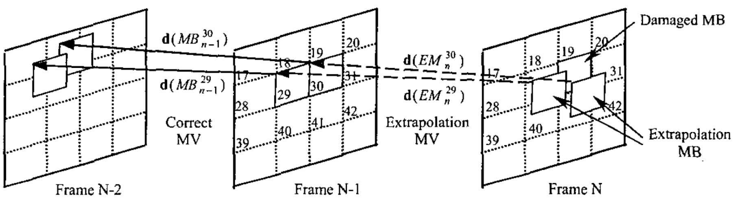


Figure 3.13: The example of motion vector extrapolation and extrapolation MBs in [7]

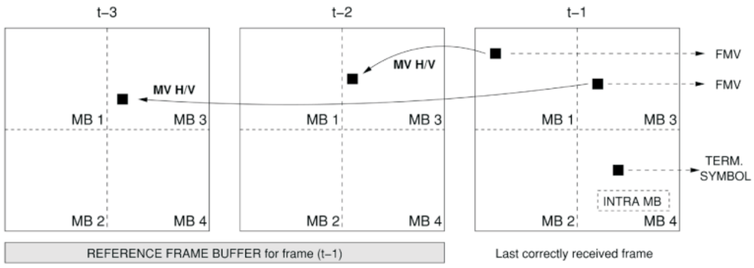


Figure 3.14: Estimation of optical flow with temporal regularization in [8] with 2 previous frames

in most cases.

Although a simple algorithm for frame loss concealment is proposed in [25], more complex algorithms that use the assumption of continuity of motion across consecutive frames are proposed for better performance in full frame concealment.

In [26], the method of motion vector extrapolation is extended to conceal the whole frame loss. The motion vector extrapolation method is applied for all the pixels in the lost frame and a motion vector is calculated for each pixel in the frame by either extrapolation or motion vector copy. The proposed method applies the motion vector copy algorithm if the extrapolation is unsuccessful for a number of pixels greater than a threshold. The method is also applied bi-directionally, averaging the calculated motion vectors of backward and forward estimation. It is concluded that bi-directional estimation outperforms forward or backward estimation used alone. It is also shown that pixel based method has a better performance than block based method.

Another pixel based method is proposed in [8]. The method is derived from the

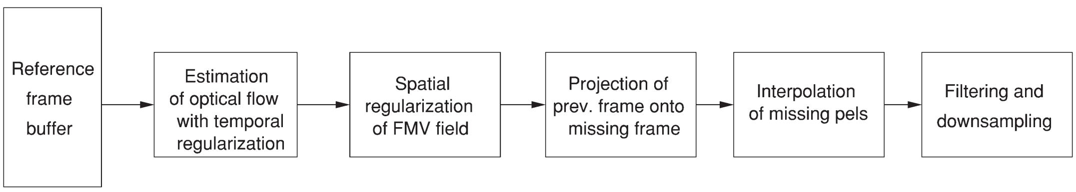


Figure 3.15: Schematic of the proposed algorithm in [8]

Figure 3.16: Schematic of the proposed algorithm in [9]

optical flow estimation from the motion vectors of multiple previous frames. The motion vectors on the same motion path over a number of previous frames are averaged to estimate the motion vector of the corresponding pixel on the trajectory of that motion in the concealed frame. The motion path, or motion trajectory, is estimated from the motion vectors of the previous frames, i.e. the motion vectors are assumed to represent the optical flow. With the estimations of the motion vectors for each pixel, a forward motion vector field is produced. The estimated motion vectors in forward motion vector field is illustrated in Figure 3.14. The vectors in this field represent the motion of each pixel in the last received frame to form the concealed frame. This motion vector field then filtered by a median filter for spatial regularization. The initial concealed frame is formed in half-pel resolution by forward motion vector field. Finally this initial estimate is interpolated, filtered by an averaging filter and downsampled to form the final estimate. The steps of the algorithm is shown in Figure 3.15. The proposed method is concluded to outperform the frame copy method in almost all cases.

The algorithm in [8], which works in the pixel domain provides good results for concealment in different cases, but it is stated to be complex to be applied in real time applications. In order to overcome this issue, an algorithm that works in block domain rather than pixels is proposed in [9]. Only a single source frame is used in this method for motion vector projection and the projected motion vectors for each pixel are averaged to estimate motion vectors for macroblocks. Although the concealed frame quality of the proposed method is reported to be lower than the results from the method in [8], the algorithm complexity is reported to be much lower and the proposed algorithm is stated to be suitable for real-time applications. The steps of the algorithm is shown in Figure 3.16.

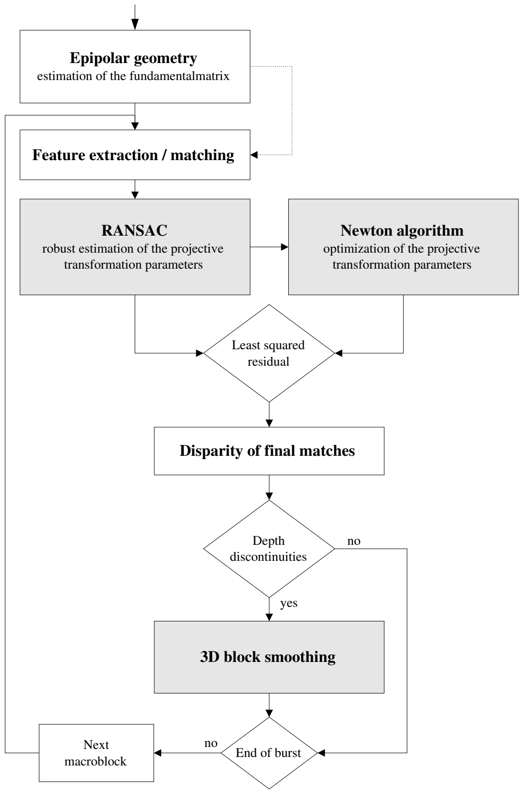


Figure 3.17: Schematic of the proposed algorithm in [10]

3.4 Concealment Algorithms For Stereo Video

Although no algorithms for concealment in multiview video is found in the literature, there are algorithms proposed for stereoscopic video and images.

In [27], [28] and [10], Sikora et al. proposed several algorithms for concealment of block losses in stereoscopic images. [27] and [10] proposes estimation of epipolar geometry and feature matching to produce a depth field (or disparity for stereoscopic videos). This depth field is then used to conceal the lost blocks. The flowchart of the proposed algorithm can be seen in Figure 3.17. It is concluded that some operations such as smoothing gives good results in objective quality, but does not effect the subjective quality. It is also stated that PSNR gives a good measure in most of the cases although comparisons should also take into account the subjective evaluation results. The presented results indicate that stereoscopic methods outperform the monoscopic methods. In [28], a hybrid approach that combines the monoscopic concealment and stereoscopic concealment is proposed. The results in this work states that depth perception is much less effected with the proposed algorithm compared to the monoscopic methods.

A full frame loss algorithm for stereoscopic video is proposed in [29]. The algorithm is implemented on a H.264 based stereoscopic codec presented in [30]. In this study the full frame loss concealment algorithm proposed in [26] is extended to stereoscopic video. The motion compensated blocks in the stereoscopic video are concealed as in the monoscopic algorithm whereas the disparity compensated blocks are concealed using the disparity vector of the block at the same position in the previous frame. The proposed method conceals the blocks in the lost frame in different groups to improve the performance. The results show that the proposed algorithm outperforms the monoscopic method proposed in [30].

CHAPTER 4

THE PROPOSED LOSS CONCEALMENT ALGORITHMS

Although there are a number of methods for the full frame loss concealment in monoscopic video in literature, the concealment of multiview video is a field that is to be investigated. The basic monoscopic methods such as frame copy can be applied on multiview video in case of frame loss. In order to improve the performance of the concealment, methods presented in [8] and [9] that use the assumption of continuity of optical flow and motion vector projection can also be applied. However in the coders that use temporal and inter-view compensation together, such as MMRG Multiview Codec and H.264 Multiview Extension, the temporal motion vector projection methods may have a degraded performance since they can not project the disparity (inter-view) vectors properly. Therefore methods that utilize both temporal and inter-view vectors via projection may perform better. In addition to better performance in the objective quality of the concealed images, the depth perception may also be improved since the depth information estimated by inter-view vectors are used in the concealment process.

In order to utilize the inter-view vectors in a basic codec structure with only I and P-frames as in MMRG Multiview Codec, two concealment algorithms that use projection of both temporal and inter-view vectors are proposed. For a more complicated reference structure with hierarchical B frames as in H.264 Multiview Extension, two algorithms that use the inter-view references in both anchor and non-anchor pictures are proposed. The use of inter-view vectors in the previous frames are prioritized over motion vectors in all proposed methods in order to investigate the improvement of the use of inter-view vectors.

4.1 The Proposed Algorithms on MMRG Multiview Codec

The full frame loss concealment algorithms such as the algorithm proposed in [8] use the assumption of the continuity of the optical flow through the consecutive frames of video. The motion vectors are assumed to represent the optical flow and the vectors of the previous frame are used to estimate the vectors of the lost frame with the assumption of linear motion. The projection of motion vectors are explained in [8] and [9]. The methods proposed in [8] and [9] will be called as Concealment Algorithm on Pixels (CAp) and Concealment Algorithm on Blocks (CAb) respectively in the following sections of this thesis.

4.1.1 Temporal Projection

Temporal projection of motion vectors is the projection among the frames in the same view. The projection is applied to the motion vectors of the previous frame and the vectors are projected onto the concealed frame. The projection in temporal direction is done as:

$$mv_{x,n,k}(i - mv_{x,n-s,k}(i,j), j - mv_{y,n-s,k}(i,j)) \\ = (s/RF_{n-s,k}(i,j)) \times mv_{x,n-s,k}(i,j)$$

$$mv_{y,n,k}(i - mv_{y,n-s,k}(i,j), j - mv_{y,n-s,k}(i,j)) \\ = (s/RF_{n-s,k}(i,j)) \times mv_{y,n-s,k}(i,j)$$

where $mv_{x,n,k}(i,j)$ and $mv_{y,n,k}(i,j)$ represent the horizontal and vertical components of the motion vector at the pixel position (i,j) of the frame at kth view at frame number n respectively. $RF_{n,k}(i,j)$ represents the frame number of the reference of the motion vector at pixel position (i,j) of the frame at kth view at frame number n respectively. Note that the term $(s/RF_{n-s,k}(i,j))$ is the scaling factor to normalize the frame gaps. The projection in temporal direction is illustrated with P1 in Figure 4.1. The temporal projection is the only projection method used in CAp and CAb algorithms.

In the case of multiview coding with inter-view compensation in addition to motion compensation, the inter-view references (spatial references or the disparity vectors) should also be projected in addition to motion vector projection. The projection of disparity vectors (inter-view vectors) are possible with two methods: hybrid projection and spatial projection.

4.1.2 Hybrid Projection

The hybrid projection uses both temporal and spatial (inter-view) references. In addition to assumptions regarding optical flow and motion vectors for temporal projection, preservation of depth of moving objects is also assumed. The hybrid projection is used on the inter-view references of the previous frame to the lost frame. The hybrid projection is done as:

$$MV_x = (s/RF_{n-s,k-1}(i,j)) \times mv_{x,n-s,k-1}(i + d_{x,n-s,k}(i,j), j + d_{y,n-s,k}(i,j))$$
$$MV_y = (s/RF_{n-s,k-1}(i,j)) \times mv_{y,n-s,k-1}(i + d_{x,n-s,k}(i,j), j + d_{y,n-s,k}(i,j))$$
$$d_{x,n,k}(i - MV_x, j - MV_y) = d_{x,n-s,k}(i,j)$$
$$d_{y,n,k}(i - MV_x, j - MV_y) = d_{y,n-s,k}(i,j)$$

where $mv_{x,n,k}(i,j)$ and $mv_{y,n,k}(i,j)$ represent the horizontal and vertical components of the motion vector at the pixel position (i,j) of the frame at kth view at frame number n respectively. And $d_{x,n,k}(i,j)$ and $d_{y,n,k}(i,j)$ represent the horizontal and vertical components of the disparity vector at the pixel position (i,j) of the frame at kth view at frame number n respectively. $RF_{n,k}(i,j)$ represents the frame number of the reference of the motion vector at pixel position (i,j) of the frame at

kth view at frame number n respectively. The hybrid projection is illustrated with P2 in Figure 4.1. In case of two cameras (stereoscopic video), only temporal and hybrid projection can be used.

4.1.3 Spatial Projection

In the case of multiview video with more than two views, references between neighboring views can be projected by spatial projection. The assumption in the spatial projection is that the depth of the objects in the scene is the same with respect to each camera. Therefore disparity for the same object between the cameras are assumed to be the same. The spatial projection is done as:

$$d_{x,n,k}(i - d_{x,n,k-v}(i,j), j - d_{y,n,k-v}(i,j)) = (v/RV_{n,k-1}(i,j)) \times d_{x,n,k-v}(i,j)$$
$$d_{y,n,k}(i - d_{y,n,k-v}(i,j), j - d_{y,n,k-v}(i,j)) = (v/RV_{n,k-1}(i,j)) \times d_{y,n,k-v}(i,j)$$

where $d_{x,n,k}(i,j)$ and $d_{y,n,k}(i,j)$ represent the horizontal and vertical components of the disparity vector at the pixel position (i,j) of the frame at kth view at frame number n respectively. $RV_{n,k}(i,j)$ represents the view number of the reference of the disparity vector at pixel position (i,j) of the frame at kth view at frame number n respectively. The spatial projection is illustrated with P3 in Figure 4.1.

It should be noted that the vectors can not be projected if any of the blocks used in the projection operation is an intra block. Therefore the number of intra blocks is important for the performance of the projection algorithms. Also if the projected vectors points to the outside of the frame boundaries the projection becomes invalid.

4.1.4 Description of the Proposed Algorithms

The proposed SPpix (Spatial Concealment on P-pictures on pixels) algorithm can be described in six steps. The algorithm starts with the selection of the source frame (or source frames in case of more than two views). The source frame is the frame vectors of which are to be projected onto the lost frame. The source frame is selected as the

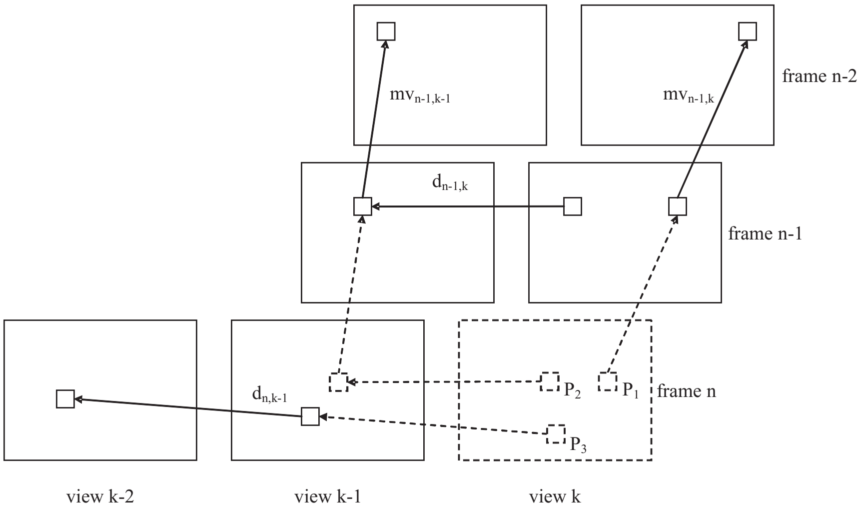
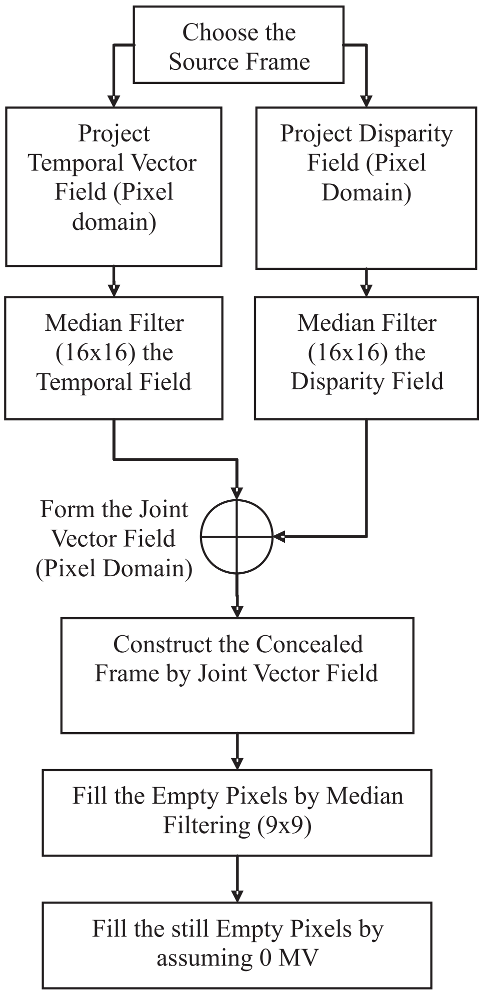
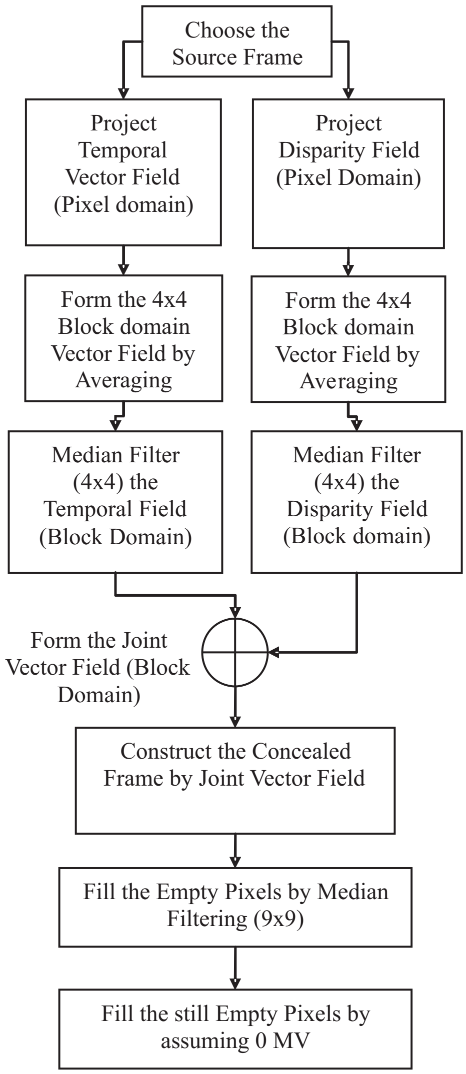


Figure 4.1: Illustration of projection methods used in MMRG Multiview Codec. P1 represents the temporal projection, P2 represents the hybrid projection and P3 represents the spatial (inter - view projection).

closest frame in the same view. If the number of views is more than two, i.e. spatial projection is available, closest frame with the same frame number is also selected as the second source frame for spatial projections. The vectors of the source frame(s) are then projected onto the lost frame to form a vector field in pixel domain, i.e. a motion vector for each pixel of the source frame is projected. The temporal projection results in temporal vector field and the hybrid and spatial projections result in the disparity vector field. After the projection step, these two vector fields are then median filtered by a kernel size of 16x16, which is proposed to be a suitable size in [8], in order to regularize the vector field for continuity. These two vector fields are then merged to form the joint vector field which will be used to reconstruct the concealed frame. When joining the vector fields, the empty regions in the spatial vector field is filled with temporal vector field. The construction of the concealed picture is done in half-pel resolution. Reconstruction in higher resolutions are reported to be unnecessary in [9] since it does not provide a significant improvement in the performance while increasing the complexity. After the reconstruction, some portions of the concealed image may still be empty if no motion vectors are projected upon the



(a) Block diagram of SPpix algorithm



(b) Block diagram of SPblk algorithm

Figure 4.2: The block diagrams of SPpix and SPblk algorithms

pixels. Therefore the reconstructed image is median filtered by a kernel size of 9x9 to fill the empty regions in the concealed image. If there still are empty pixels in the concealed image after median filtering, these pixels are filled by assuming 0 motion vector. Finally the image is downsampled to the original resolution.

The steps of median filtering the vector fields with a kernel size of 16x16 is an operation with high complexity and may not be applicable in real-time applications. Therefore the SPblk (Spatial Concealment on P-pictures on blocks) algorithm with lower complexity is also proposed. In SPblk algorithm each of the vector fields in pixel domain are averaged to provide a motion vector to a 16x16 block of pixels in the half-pel resolution picture. After this averaging, the vector field sizes are decreased to 1/16 of the vector fields in the SPpix algorithm. In addition to the decrease in the size of the vector fields, the median filter kernel size for vector fields is also decreased to 4x4 to provide fast filtering of the vector fields. The other steps of the SPblk algorithm is the same as SPpix algorithm.

The block diagrams of the SPpix and SPblk algorithms can be seen in Figure 4.2.

4.2 The Proposed Algorithms on H.264 Multiview Extension

The main difference between the MMRG Multiview Codec and H.264 Multiview Extension is in their referencing structure and the use of hierarchical B-pictures. Due to these differences the projection methods are also modified accordingly in the H.264 Multiview Codec. Another important difference is the anchor pictures in the H.264 Multiview Extension. The frames of each view in the beginning of each GOP are called anchor frame, and these frames do not have any temporal references. Therefore all the blocks in these frames are encoded with either intra mode or inter-view compensation. These references can be used in favor of the full frame concealment to make use of all available inter-view compensation information. The bi-directionality of the B-pictures should also be taken into account. Since the B-pictures have temporal references in both backward and forward directions, both of these references can be used for bi-directional estimation.

The temporal, hybrid and spatial projection methods are also available in H.264 Multiview Extension. The temporal projection structures for hierarchical B-pictures are described in [26] in detail and are applicable to all the views. The hybrid projection can only be used on the views that use inter-view compensation on non-anchor pictures (views 1, 3 and 5 in Figure 2.7) when general scheme of the H.264 Multiview Extension is used. The spatial projection is applicable only to the anchor pictures both in general and simplified schemes.

In case of the losses of the B pictures at different levels, the available frames to the decoder are also different. For instance when concealing a B2 frame loss, the available, i.e. already decoded, frames are only I, P and B1 frames. Therefore the concealment procedures differ for different types of frames.

The proposed SHBpix (Spatial Concealment on Hierarchical B-pictures on pixels) algorithm uses the spatial references in the anchor frames as well as the spatial references in the non-anchor pictures if available. The difference of the SHBpix algorithm from the SPpix algorithm is that the forming of the vector fields is different and only a spatial vector field (disparity field) is formed while no temporal field is used. Instead forward and backward disparity fields are formed by the forward and backward references of B-pictures and these two fields are averaged to form an intermediate bi-directionally estimated disparity field. The other steps of the algorithm is the same as SPpix algorithm as it can be seen in Figure 4.7(a) except in the final step, instead of a zero vector, average value of the disparity field is assumed for the empty pixels of the concealed picture. This is due to the nature of disparity vectors that unlike motion vectors, the disparity vectors are rarely zero since all the objects in the scene have a finite depth and disparity except the objects at infinity (or very far away)in parallel camera configurations.

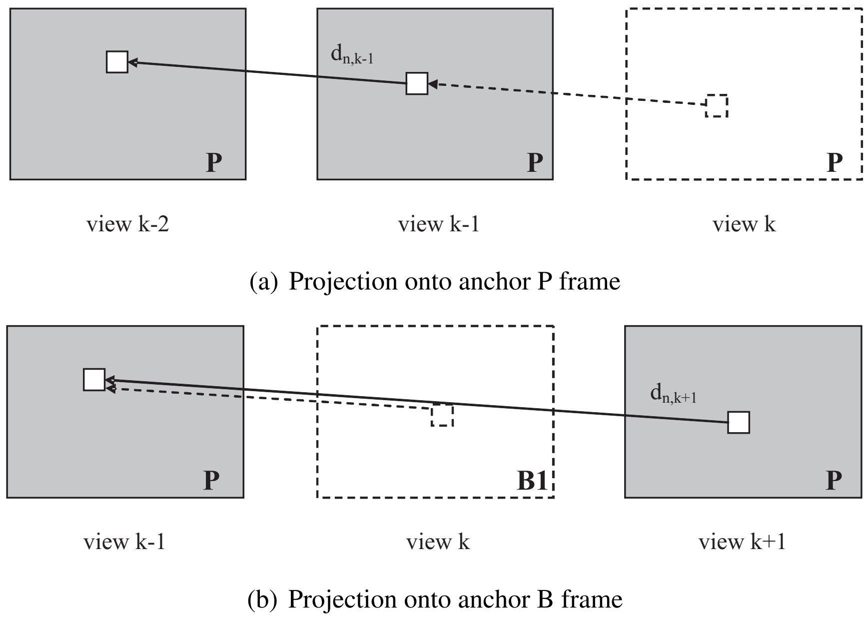


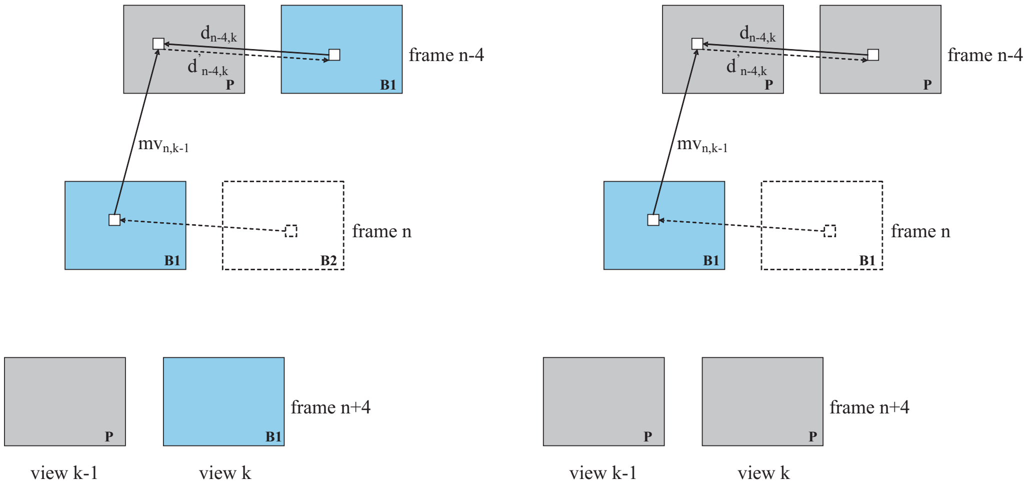
Figure 4.3: Illustrations of projections for the concealment of anchor pictures in SHBpix algorithm

4.2.1 Projections to Form the Disparity Field in Anchor Pictures with SHBpix

The concealment of anchor pictures are done only with information from other anchor pictures with the same frame number, because the previously decoded anchor pictures with different frame number does not have any relation or common reference to the anchor picture to be concealed. As a result the vectors are projected only via spatial projection. Note that the anchor pictures only have inter-view references. The projections to form the disparity field of concealment are illustrated in Figure 4.3.

4.2.2 Projections to Form the Disparity Field in First Level Non-Anchor B Pictures with SHBpix

The first level non-anchor B pictures (B2 frames in the views with inter-view references in general scheme and B1 frames in simplified scheme of H.264 Multiview Extension) are concealed with the help of the disparity vectors in the anchor pictures before and after the concealed picture. Before the projection, the disparity vectors



(a) The backward projection of the anchor disparity vectors in general scheme

(b) The backward projection of the anchor disparity vectors in simplified scheme

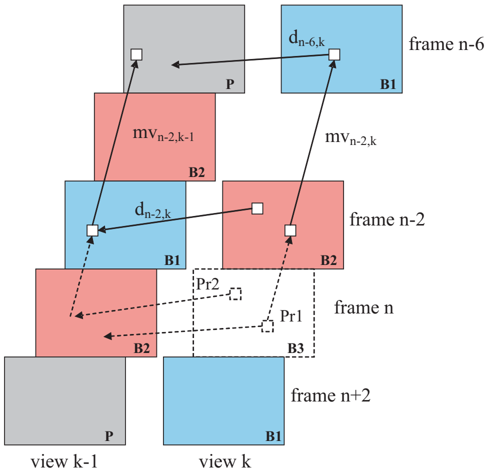
Figure 4.4: The backward projection of the anchor disparity vectors for the concealment of first level non-anchor B-Pictures in different schemes of H.264 Multiview Extension

of the anchor picture (both the pictures in the backward and forward direction) are inverted to form the inverse disparity field of anchor pictures as:

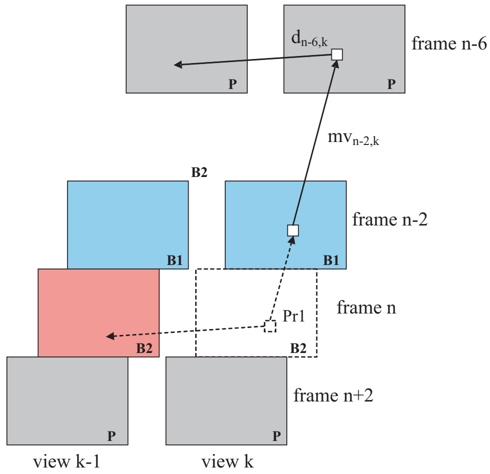
$$d'_{x,n,k-1}(i + d_{x,n,k}(i,j), j - d_{y,n,k}(i,j)) = (-1/RV_{n,k}(i,j)) \times d_{x,n,k-1}(i,j)$$
$$d'_{y,n,k-1}(i + d_{y,n,k}(i,j), j - d_{y,n,k}(i,j)) = (-1/RV_{n,k}(i,j)) \times d_{y,n,k-1}(i,j)$$

where $d'_{x,n,k}(i,j)$ and $d'_{y,n,k}(i,j)$ are the horizontal and vertical components of the inverted disparity vector at the pixel position (i,j) of the frame at kth view at frame number n respectively. Similarly $d_{x,n,k}(i,j)$ and $d_{y,n,k}(i,j)$ represent the horizontal and vertical components of the disparity vector at the pixel position (i,j) of the frame at kth view at frame number n respectively. $RV_{n,k}(i,j)$ represents the view number of the reference of the disparity vector at pixel position (i,j) of the frame at kth view at frame number n respectively.

The inverted disparity vectors are then projected to form the forward and backward disparity fields. The backward projection of the inverted disparity vectors is illustrated in Figure 4.4 and the forward projection is also done similarly with the

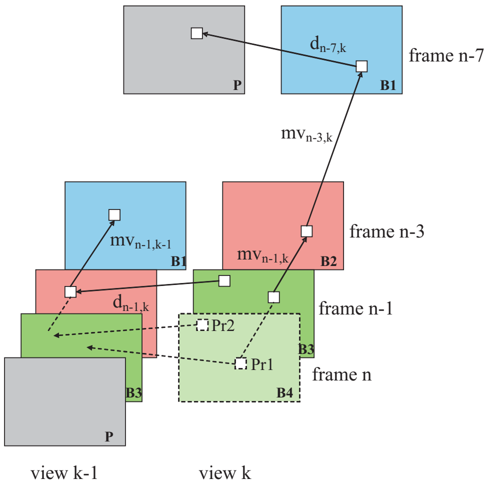


(a) The backward projection of the anchor disparity vectors in general scheme

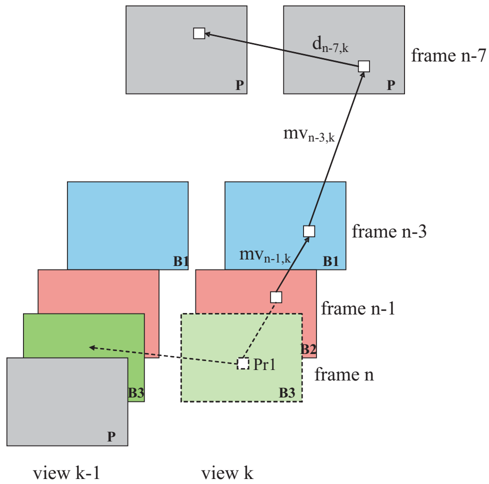


(b) The backward projection of the anchor disparity vectors in simplified scheme

Figure 4.5: The backward projection of the anchor disparity vectors for the concealment of second level non-anchor B-Pictures in different schemes of H.264 Multiview Extension



(a) The backward projection of the anchor disparity vectors in general scheme



(b) The backward projection of the anchor disparity vectors in simplified scheme

Figure 4.6: The backward projection of the anchor disparity vectors for the concealment of third level non-anchor B-Pictures in different schemes of H.264 Multiview Extension

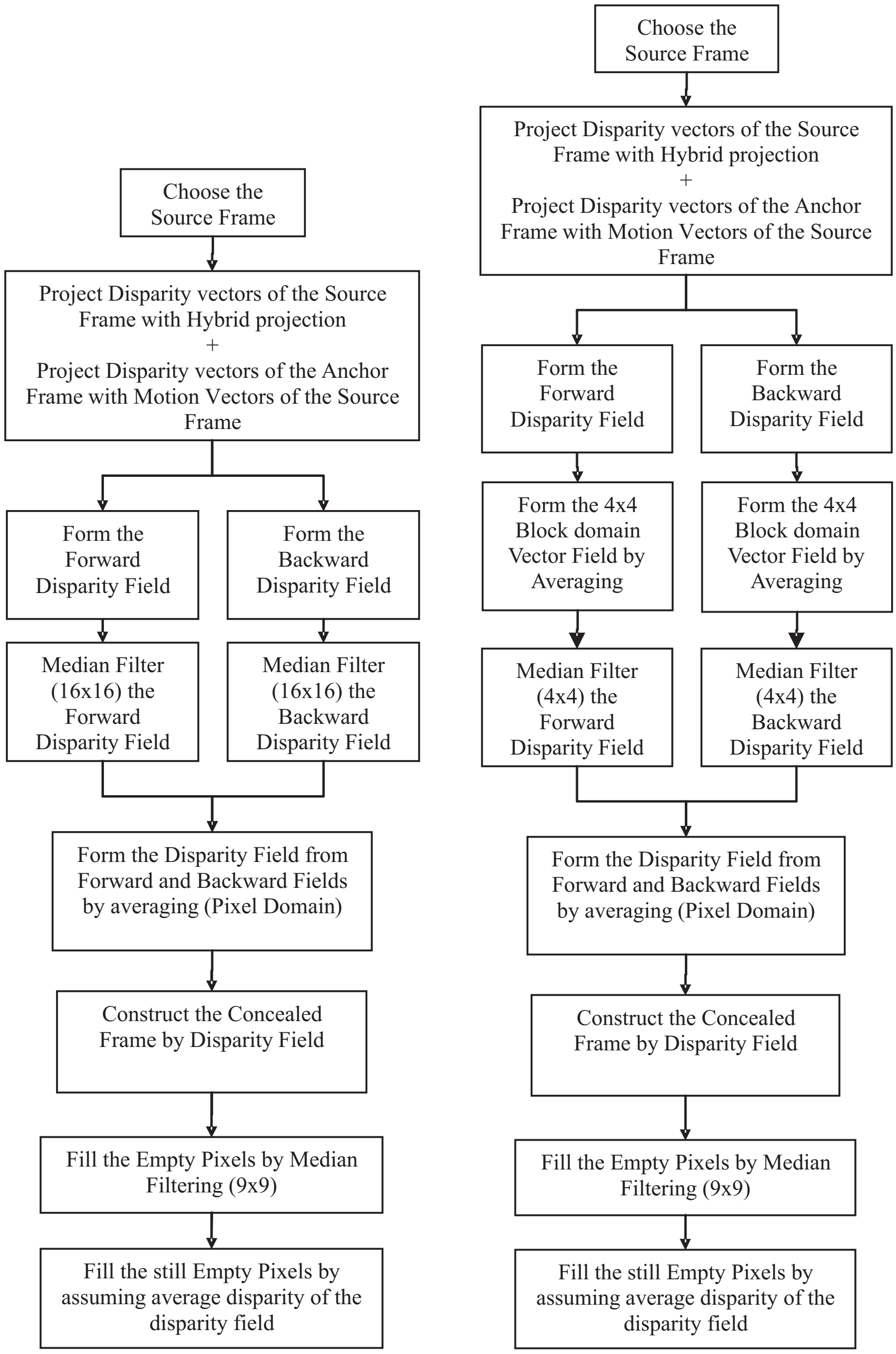
forward references of the source frame. The pixel based forward and backward disparity fields are then averaged to form the disparity field for concealment with SHBpix algorithm.

4.2.3 Projections to Form the Disparity Field in Second and Higher Level Non-Anchor B Pictures with SHBpix

The procedure for the second level (B3 frames in the views with inter-view references in general scheme and B2 frames in simplified scheme of H.264 Multiview Extension) or higher level non-anchor B-pictures are almost the same except the number of levels of projection.After choosing the source frame, which is selected as the temporally closest frame to the lost frame within the same view, the disparity vectors of the source frame are projected by hybrid projection as in SPpix algorithm (illustrated with Pr2 in Figures Figure 4.5(a) and Figure 4.6(a)). But the temporal vectors are not used to form a temporal vector field but to project the disparity vectors of the anchor frames onto the lost frame. The backward projection, i.e. projection with backward vectors of the B frames, of anchor disparity vectors onto the lost frame is illustrated with Pr1 in Figures Figure 4.5(a), Figure 4.5(b), Figure 4.6(a) and Figure 4.6(b). It should be noted that as the level of the concealed frame increases, the number of temporal steps to the anchor view from the source frame may increase and this could reduce the reliability of the operation. With backward projection of the anchor frame inter-view references and also the hybrid projection of non-anchor frame inter-view references, the backward disparity field is formed. Similarly forward disparity field is formed by forward projections. The backward and forward disparity fields are then averaged to form the disparity field for the concealment.

4.2.4 SHBblk Algorithm

As in the case with SPpix, the SHBpix algorithm also has a high complexity for real-time applications. Therefore similar to SPblk, SHblk algorithm is proposed to reduce the complexity by working in the block domain rather than pixel domain for vector



(a) Block diagram of SHBpix algorithm (b) Block diagram of SHBblk algorithm

Figure 4.7: The block diagrams of SHBpix and SHBblk algorithms

field processing. After the projection of motion and disparity vectors and forming the forward and backward disparity fields as in SHBpix algorithm, the 4x4 disparity vector blocks are averaged to reduce into one disparity vector to form a disparity field with a size of 1/16 of the disparity field in SHBpix algorithm. The median filter kernel size is also reduced to 4x4 to provide low complexity. After forming block domain disparity field and median filtering, the concealed picture is formed in half-pel with this vector field as in the case of SHBpix.

CHAPTER 5

EXPERIMENTAL RESULTS

The algorithms on multiview video must be tested with respect to both objective and subjective measures in order to evaluate the perceptional depth quality of these videos. Due to the lack of a proper subjective evaluation setup for multiview video, the experiments on the proposed full frame loss concealment algorithms are conducted for number of views equal to two, i.e. stereoscopic case. Only one multiview video is tested to evaluate the objective quality of the proposed concealment algorithms in case of high camera distance. Since only stereoscopic videos are tested, the proposed algorithms are applied on only one of the views and the other view is encoded without inter-view references as it can be seen from stereoscopic reference structures in Figures Figure 2.6, Figure 2.7 and Figure 2.8.

5.1 Objective Quality Comparison Procedure

The comparisons of objective quality are made with peak signal-to-noise ratio (PSNR) metric which is commonly used in video and image coding literature. The PSNR of two 8-bit digital images is calculated with:

$$PSNR = 10 \times \log(255^2 / \sum_{i=1}^{w} \sum_{j=1}^{h} (im1_{i,j} - im2_{i,j})^2) \tag{5.1}$$

where w is image width, h is image height, $im1_{i,j}$ is pixel value of first image at position (i, j) and $im2_{i,j}$ is pixel value of second image at position (i, j). The unit for PSNR is in decibels (dBs).

Table 5.1: The stereoscopic videos used in the experiments. The Breakdancers video is a multiview video, only first two views of this video is used for the tests

Video Name	Resolution (Width , Height)
Xmas	(640,480)
Rena	(640,480)
AkkoKayo	(640,480)
Botanical	(960,540)
Soccer	(720,480)
Train	(720,576)
Balloons	(720,480)
Heidelberg	(640,480)
Mouldpenny	(640,480)
Breakdancers	(640,384)

5.1.1 Comparisons of Methods on MMRG Multiview Codec

In order to test the proposed methods on MMRG Multiview Codec (SPpix and SPblk methods), first 72 frames of the stereoscopic videos in Table 5.1 are encoded with MMRG Multiview Codec with number of views equal to 2, number of reference frames equal to 5, search range equal to 32, reference mode equal to 1 (the reference structure is as in Figure 2.6). To compare the performances of SPpix and SPblk methods with CAp and CAb methods, right views of these videos are also encoded as simulcast with standard H.264 codec with same options as MMRG Multiview Codec. For the loss of each frame in the right view, the lost frame is concealed with frame-copy, SPpix, CAp, SPblk and CAb methods and the completion times (T_{tot}) as well as just the motion vector projection times (T_{mv}) are recorded. The PSNR between each concealed picture and the correctly decoded picture is also calculated for comparison.

The average PSNR results for each video and for each concealment method are shown in Table 5.2. It is clear from the results that objective quality of the SPpix algorithm is better than other methods in most of the videos. In the cases where SPpix is outperformed, the difference in PSNR is less than 1 dB. It should be also noted that the performance of SPblk method is superior to both CAb and Frame Copy methods in all videos except the Heidelberg video which has very low motion characteristics and Frame Copy gives very high PSNR results. In some of the cases SPblk method even outperformed CAp method which has a higher complexity than SPblk due to working in pixels.

Table 5.2: The average PSNR values of the concealment results of each method for each frame in the video sequences. The best result for each video is shown in bold. The results are in dBs.

Video Name	Frame Copy	SPpix	CAp	SPblk	CAb
Xmas	20.199	**30.812**	21.044	30.417	19.798
Rena	30.748	**35.126**	33.218	32.498	29.645
AkkoKayo	22.627	30.281	**30.343**	24.684	22.722
Botanical	27.577	**35.273**	28.955	34.410	25.985
Soccer	24.374	**31.597**	21.044	27.341	23.971
Train	25.372	**26.623**	26.466	25.435	24.626
Balloons	24.976	**28.384**	27.469	26.431	25.662
Heidelberg	**30.727**	29.805	28.541	26.895	27.006
Mouldpenny	26.599	28.993	**29.257**	27.060	26.227
Breakdancers	24.869	**27.390**	25.488	26.823	24.785

The statistics of encoding mode of the blocks in the frames of the encoded videos are shown in Table 5.3. It can be observed that as the number of inter-view compensated blocks increase in the frames the performance of the SPpix algorithm increases especially with respect to Frame Copy. SPblk algorithm also behaves similarly to SPpix algorithm. The performance of all the projection algorithms are also inversely proportional to the percentage of the intra coded blocks in the encoded videos. This is reasonable since intra coded blocks are estimated from the neighboring motion vector and pixel information and increase in the number of intra blocks increase the probability of wrong estimations.

Table 5.3: The average percentages of intra coded blocks and spatially coded (inter-view compensated) blocks in the frames of each video when encoded with MMRG Multiview Codec (MMRGMC). The average percentage of intra coded blocks in the frames of each video when encoded with H.264 are also presented.

Video Name	Intra Blocks in MMRGMC (%)	Spatial Blocks in MMRGMC (%)	Intra Blocks in H.264 (%)
Xmas	2	34	7
Rena	7	12	10
AkkoKayo	2	12	4
Botanical	0	35	7
Soccer	9	11	10
Train	2	9	5
Balloons	6	8	9
Heidelberg	0	5	2
Mouldpenny	2	7	3
Breakdancers	9	12	14

The algorithm complexity can be observed via the total time for the concealment with each algorithm which is shown in Table 5.4 for each video. The algorithms are run on a AMD Athlon XP 3200+ computer with 768 Megabyte memory. It can be observed in Table 5.4 that, block based algorithms reduced the total concealment time to about 10% of the pixel based algorithms which is a significant improvement over complexity. This reduction is mainly due to lowering the median filtering complexity in the block based algorithms which can also be observed via motion vector projection times (T_{mv}) in Table 5.4. In order to observe the algorithm completion time in different resolutions some of the videos are also downsampled to 320 x 240 resolution and concealed after being encoded as in the original size. The results for algorithm times at 320 x 240 resolution can be seen in Table 5.5. At this resolution the algorithm completion times are lower than 1 second and for block based algorithms it is lower than 70 milliseconds. As a result the SPblk algorithm is suitable for real-time applications and the SPpix can also be used with the use of a suitable buffer for decoding.

Table 5.4: The average algorithm completion times (T_{tot}) and motion vector projection times (T_{mv}) in milliseconds for the algorithms for each tested video.

Video Name	SPpix (T_{tot})	CAp (T_{tot})	SPblk (T_{tot})	CAb (T_{tot})	SPpix (T_{mv})	CAp (T_{mv})	SPblk (T_{mv})	CAb (T_{mv})
Xmas	3878	3387	346	338	3638	2805	67	52
Rena	3849	3017	395	373	3592	2571	64	49
AkkoKayo	3952	3133	341	314	3716	2693	69	52
Botanical	6640	5692	618	569	6215	4659	112	92
Soccer	4387	3566	405	362	4137	3020	74	59
Train	5291	4225	493	471	4963	3558	88	69
Balloons	4483	3654	390	373	4235	3070	76	59
Heidelberg	3936	2885	294	283	3724	2608	67	49
Mouldpenny	4031	3064	307	289	3812	2714	68	53
Breakdancers	3110	2424	408	374	2849	2013	51	40

Table 5.5: The average algorithm completion times (T_{tot}) and motion vector projection times (T_{mv}) in milliseconds for the algorithms for some of the tested videos at 320 x 240 resolution.

Video Name	SPpix (T_{tot})	CAp (T_{tot})	SPblk (T_{tot})	CAb (T_{tot})	SPpix (T_{mv})	CAp (T_{mv})	SPblk (T_{mv})	CAb (T_{mv})
Xmas	929	838	58	59	892	705	16	15
Botanical	929	803	58	59	888	675	16	12
Balloons	976	788	68	71	936	681	16	14
Train	963	778	65	64	919	660	16	12
Soccer	959	776	71	65	911	666	15	12

5.1.2 Comparisons of Methods on H.264 Multiview Extension

In order the test the performances of the algorithms SHMpix and SHBblk and compare with Frame Copy, CAp and CAb algorthms the test videos are encoded with the general reference structure of H.264 Multiview Extension which is shown in Figure 2.7. In addition to this, the videos are also encoded with the simplified reference structure seen in Figure 2.8 since the temporal methods are expected to work better on this reference structure due to less number of inter-view predicted pictures.

The frame copy algorithm in H.264 Multiview Extension has a difference from the one applied to MMRGMC in the sense that the frame to be copied for concealment is not always the previous frame to the frame to be concealed. This is because of the hierarchical B-pictures in the H.264 Multiview Extension referencing structure. For instance when a first level non-anchor B frame is lost in the right view, the frame to be copied instead is the frame that is four frames earlier in the right view. As a result the PSNR figures are expected to be lower for lower level B frames compared to B4 frames.

The anchor frames are not considered in the tests since none of the tested concealment methods is applicable to conceal the anchor pictures for the stereoscopic case. Each frame in the encoded videos are concealed with the monoscopic and proposed methods. Since no temporal references between the frames exist while concealing the first level non-anchor B frames, CAp and CAb algorithms are not applicable for the concealment. Therefore the comparison is made only to the frame copy algorithm.

The results of average PSNR for the general reference scheme of H.264 Multiview Extension are presented in Table 5.6, Table 5.8 and Table 5.10 for first, second and third level non-anchor B frames respectively. Similarly the results of average PSNR for the simplified reference scheme of H.264 Multiview Extension are presented in Table 5.7, Table 5.9 and Table 5.11.

For the concealment of first level non-anchor B pictures, the proposed methods are both superior to Frame Copy both for general and simplified schemes except the case with Botanical sequence. Since this video has a very stationary scene with very slow motion, the Frame Copy PSNR is very high especially for higher level B frames. In such a case of very slow or no motion, the methods that use the disparity primarily for concealment fail to outperform temporal methods such as frame copy. Other than Botanical sequence, the concealment of B2 frames with proposed SHBpix and SHBblk methods are successful and superior compared to Frame Copy.

In the concealment of second level non-anchor B frames, the proposed methods are superior to temporal methods in six of the tested sequences both in the general and simplified schemes. In the sequences Botanical, Train, Balloons and Heidelberg the performances of the temporal methods are better. All of these videos have a common characteristic that they have a stationary scene with slowly moving objects. For instance the beginning of the Train sequence has high motion where all the algorithms have similar performances whereas the motion slows down towards the end of the sequence and the temporal methods gain on the proposed methods (Figure 5.5 and Figure 5.6). Similarly in the Balloons sequence which has slow motion in the beginning temporal methods have higher performances, whereas proposed methods start to outperform the temporal methods towards the end of the sequence where both objects and the camera have higher motion (Figure 5.3 and Figure 5.4). The Heidelberg sequence is also a very stationary sequence with high frame rate except at the

Table 5.6: The average PSNR values of the concealment results of each method for each first level non-anchor B frame in the video sequences when encoded with general scheme of H.264 Multiview Extension. The best result for each video is shown in bold. The results are in dBs.

Video Name	Frame Copy	SHBpix	SHBblk
Xmas	17.800	28.935	**30.383**
Rena	25.398	**28.482**	28.394
AkkoKayo	16.821	26.678	**26.728**
Botanical	**26.801**	23.962	25.112
Soccer	20.722	27.828	**28.913**
Train	20.594	22.063	**22.380**
Balloons	20.344	23.141	**23.579**
Heidelberg	21.010	22.518	**24.703**
Mouldpenny	20.342	**25.160**	24.743
Breakdancers	22.468	**26.320**	25.350

Table 5.7: The average PSNR values of the concealment results of each method for each first level non-anchor B frame in the video sequences when encoded with simplified scheme of H.264 Multiview Extension. The best result for each video is shown in bold. The results are in dBs.

Video Name	Frame Copy	SHBpix	SHBblk
Xmas	17.801	28.696	**30.066**
Rena	25.399	**28.436**	28.353
Akkokayo	16.821	26.639	**26.694**
Botanical	**26.801**	23.884	24.997
Soccer	20.722	27.796	**28.833**
Train	20.595	22.055	**22.371**
Balloons	20.344	23.118	**23.548**
Heidelberg	21.010	22.514	**24.696**
Mouldpenny	20.342	**25.122**	24.711
Breakdancers	22.469	**26.276**	25.321

beginning (Figure 5.7 and Figure 5.8). These three sequences also have a low ratio of inter-view predicted blocks as it can be observed from Table 5.12. In the Botanical video which is a special case such that although it has a very high percentage of inter-view compensated blocks, the frame difference PSNR is so high that concealment methods can not outperform the Frame Copy algorithm. This is due to high number of zero motion vectors in the scene. In the methods SPpix and SPblk, the zero motion vectors can be used as advantage similar to CAp and CAb algorithms since these vectors are projected without any error whereas disparity vectors, which are most of the time nonzero, have errors in the projection due to multiple reasons such as quantization of vectors, non-linearity of changes in the vector field or deviations from the actual disparity in the disparity estimations made by the encoder. Therefore the methods SHBpix and SHBblk may have problems in the estimation of stationary scenes which also causes the performance degradation in Heidelberg, Train and Balloons sequences. However this disadvantage becomes an advantage when the sequences have high motion or camera motion, i.e. when more than a percentage of the scene have nonzero motion vectors.

Table 5.8: The average PSNR values of the concealment results of each method for each second level non-anchor B frame in the video sequences when encoded with general scheme of H.264 Multiview Extension. The best result for each video is shown in bold. The results are in dBs.

Video Name	Frame Copy	SHBpix	CAp	SHBblk	CAb
Xmas	19.268	**28.079**	16.307	27.572	15.794
Rena	27.394	**29.885**	27.729	28.906	27.680
Akkokayo	19.486	**26.128**	23.441	25.063	23.155
Botanical	**27.721**	23.891	16.341	24.176	15.617
Soccer	22.152	**28.532**	23.481	28.202	23.229
Train	**22.719**	21.812	22.710	21.540	22.656
Balloons	22.733	23.058	**24.683**	22.405	24.591
Heidelberg	24.970	25.232	**28.560**	25.671	28.437
Mouldpenny	23.123	24.857	23.729	**24.890**	23.588
Breakdancers	23.689	**25.702**	24.141	25.264	24.112

For first and second level non-anchor B frames, the temporal methods have disadvantages compared to third level non-anchor B frames since the temporal distance is higher in lover levels. In the case of third level non-anchor B frames on the general scheme the temporal methods become superior to the proposed methods (SHBpix and SHBblk) in sequences Akkokayo and Mouldpenny as well. In the comparisons on the simplified scheme on which temporal methods work better compared to general scheme due to the lack of spatial (inter-view) references, the methods SHBpix and SHBblk are only superior only in the Xmas sequence which is an ideal stereo-

Table 5.9: The average PSNR values of the concealment results of each method for each second level non-anchor B frame in the video sequences when encoded with simplified scheme of H.264 Multiview Extension. The best result for each video is shown in bold. The results are in dBs.

Video Name	Frame Copy	SHBpix	CAp	SHBblk	CAb
Xmas	19.268	**26.342**	19.639	26.085	19.711
Rena	27.394	**29.738**	28.635	28.954	28.688
Akkokayo	19.486	25.311	**25.562**	24.493	25.379
Botanical	**27.721**	23.601	25.986	23.798	25.471
Soccer	22.152	**27.639**	25.406	27.476	25.354
Train	22.719	21.529	23.215	21.448	**23.273**
Balloons	22.733	23.963	25.507	22.618	**25.760**
Heidelberg	24.970	25.218	28.954	25.640	**29.081**
Mouldpenny	23.123	**24.655**	24.112	24.536	24.112
Breakdancers	23.689	**25.438**	24.529	25.300	24.602

scopic sequence without any illumination difference, motion blur and with very high percentage of inter-view references and can be considered as an exception. Therefore these methods fail to outperform the temporal methods in third level non-anchor B frames in the simplified scheme. In the general scheme, SHBpix and SHBblk are successful only in the sequences with fast motion characteristics and with camera motion (Breakdancers and Rena have fast moving objects and Soccer has a camera motion).

Table 5.10: The average PSNR values of the concealment results of each method for each third level non-anchor B frame in the video sequences when encoded with general scheme of H.264 Multiview Extension. The best result for each video is shown in bold. The results are in dBs.

Video Name	Frame Copy	SHBpix	CAp	SHBblk	CAb
Xmas	20.238	30.304	17.711	**30.716**	17.236
Rena	30.560	32.615	30.942	**32.905**	30.547
Akkokayo	22.514	27.604	28.606	27.936	**28.871**
Botanical	**28.496**	24.177	16.373	25.427	15.664
Soccer	24.028	30.770	26.483	**31.282**	26.068
Train	25.110	22.936	25.670	23.621	**25.691**
Balloons	26.152	24.956	**28.707**	25.600	28.647
Heidelberg	29.728	25.346	32.093	26.128	**32.094**
Mouldpenny	26.835	25.313	**29.427**	26.240	29.217
Breakdancers	26.023	27.007	25.594	**27.129**	25.600

In the results presented in Tables Table 5.6-Table 5.11, it can be seen that the block based methods have comparable results and sometimes even better results to the pixel based methods unlike the methods in MMRGMC. When the hierarchical B-pictures are used in temporal prediction, the projection in the temporal direction

Table 5.11: The average PSNR values of the concealment results of each method for each third level non-anchor B frame in the video sequences when encoded with simplified scheme of H.264 Multiview Extension. The best result for each video is shown in bold. The results are in dBs

Video Name	Frame Copy	SHBpix	CAp	SHBblk	CAb
Xmas	20.238	**28.258**	21.394	27.772	21.452
Rena	30.560	31.392	33.384	31.759	**33.445**
Akkokayo	22.514	25.160	30.114	25.564	**30.658**
Botanical	**28.496**	23.574	27.475	24.550	27.058
Soccer	24.028	28.962	29.103	28.626	**29.115**
Train	25.110	22.233	**26.593**	22.810	26.579
Balloons	26.152	25.274	29.354	26.363	**29.535**
Heidelberg	29.728	25.151	**31.941**	25.565	31.724
Mouldpenny	26.835	24.496	**29.809**	25.056	29.752
Breakdancers	26.023	25.512	25.916	25.636	**26.025**

Table 5.12: The average percentages of intra coded blocks and spatially coded (inter-view compensated) blocks in the backward direction (indicated B) in the frames of each video when encoded with general (indicated with G) and simplified (indicated with S) schemes of H.264 Multiview Extension. The average percentage of intra coded blocks in forward direction (indicated F) in the frames of each video when encoded with H.264 are also presented.

Video Name	Intra (G, B)	Spatial (G, B)	Intra (G, F)	Intra (S, B)	Spatial (S, B)	Intra (S, F)
Xmas	7	42	47	17	14	32
Rena	14	12	27	16	7	25
Akkokayo	13	16	30	14	10	28
Botanical	5	38	43	17	14	30
Soccer	22	15	35	25	7	32
Train	21	15	33	23	13	34
Balloons	13	14	38	17	9	35
Heidelberg	7	12	23	7	13	23
Mouldpenny	10	13	26	11	11	25
Breakdancers	20	16	36	25	9	35

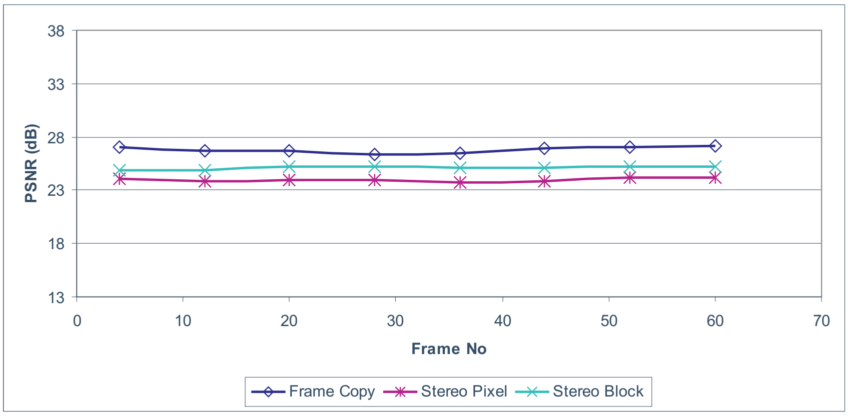
becomes less accurate and erroneous results are more likely to occur. But in the block based methods the projected motion vectors are averaged and the picture is concealed in blocks rather than pixels which may sometimes result in better concealment performances compared to pixel based methods.

The algorithm complexity can be observed via the total time for the concealment with each algorithm which is shown in Table 5.13 for each video. The algorithms are run on a AMD Athlon XP 3200+ computer with 768 Megabyte memory. In general there is about 15% increase in the completion time (T_{tot}) in SHBblk algorithm compared to SPblk presented in Table 5.4. The increase in the completion time of the SHBpix with respect to SPpix is much more significant due to greater number of median filters on the motion vector fields. But since the performance of SHBblk is very close to (and sometimes better than) SHBpix, it can be used instead. The algorithm completion time of SHBblk is not greater than 0.5 seconds and it is expected to be much smaller in the smaller resolutions just as SPblk. Therefore SHBblk is suitable for using in real-time applications.

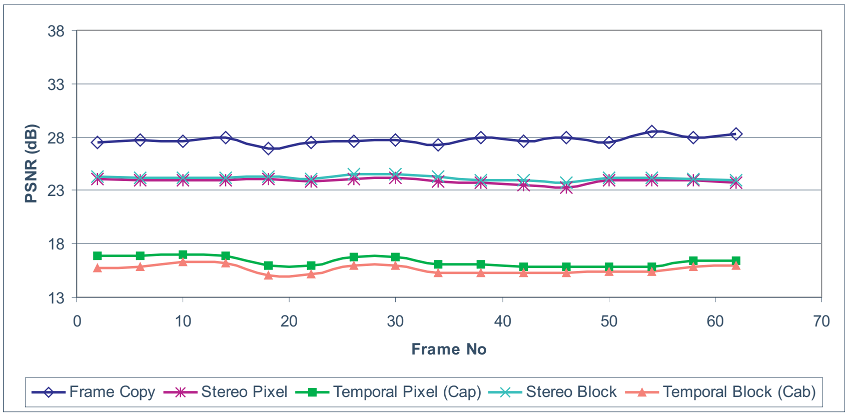
Table 5.13: The average algorithm completion times (T_{tot}) and motion vector projection times (T_{mv}) in milliseconds for the algorithms in H.264 Multiview Extension General scheme for each tested video.

Video Name	SHBpix (T_{tot})	CAp (T_{tot})	SHBblk (T_{tot})	CAb (T_{tot})	SHBpix (T_{mv})	CAp (T_{mv})	SHBblk (T_{mv})	CAb (T_{mv})
Xmas	8414	2525	402	364	8132	1749	127	24
Rena	8461	2384	461	397	8109	2070	130	30
Akkokayo	9053	2424	419	385	8703	2091	145	31
Botanical	10128	2753	566	489	9712	2008	157	28
Soccer	8892	2705	477	415	8540	2323	134	34
Train	11933	3240	717	637	11336	2727	181	40
Balloons	9686	2716	470	421	9316	2397	153	37
Heidelberg	10192	2351	379	239	9900	2168	171	30
Mouldpenny	9520	2432	388	304	9205	2199	158	32
Breakdancers	6574	1900	421	375	6232	1618	106	24

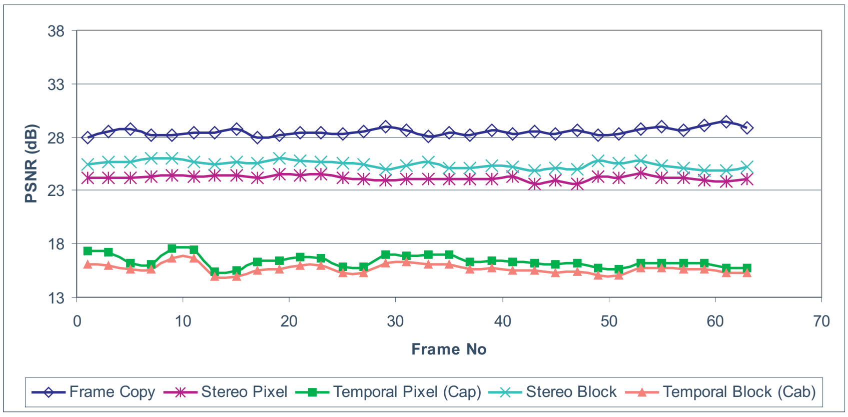
The average concealment results in PSNR and average encoding statistics of the videos can be found in Tables Table 5.6-Table 5.12.



(a) Results of B2 pictures

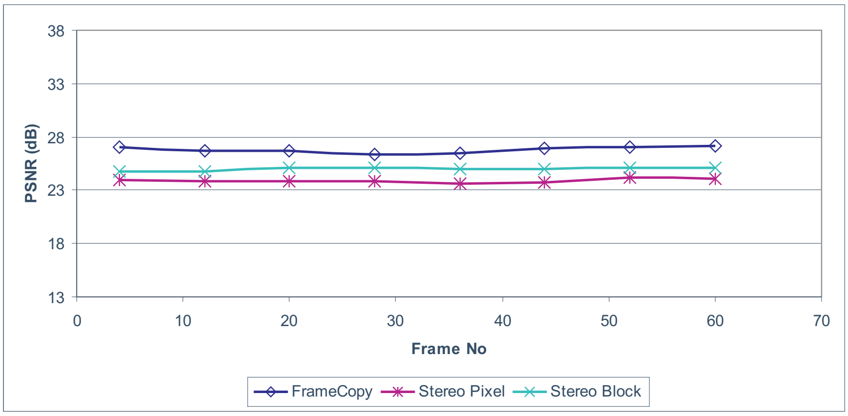


(b) Results of B3 pictures

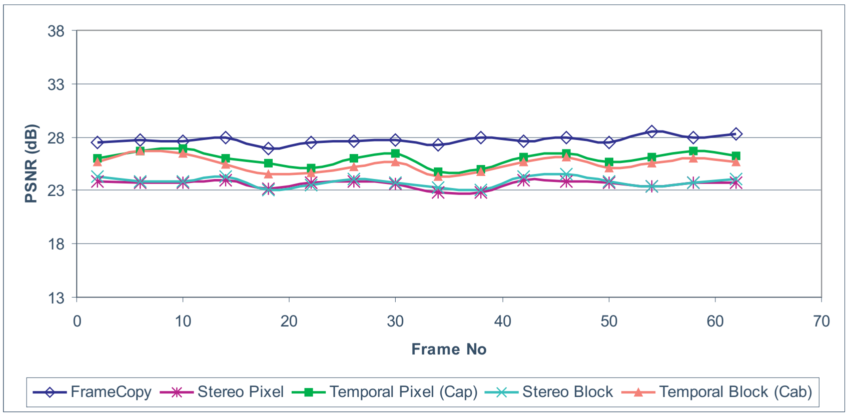


(c) Results of B4 pictures

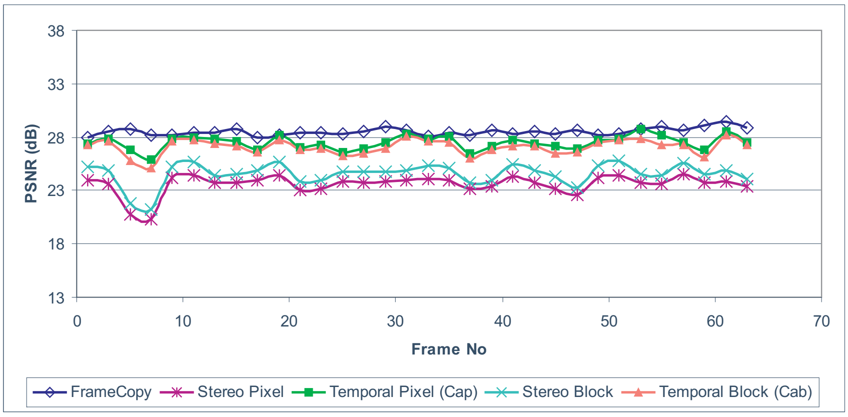
Figure 5.1: Concealment results of Botanical sequence when encoded with General referencing structure of H.264 Multiview Extension (Stereo Pixel = SHBpix, Stereo Block = SHBblk)



(a) Results of B2 pictures

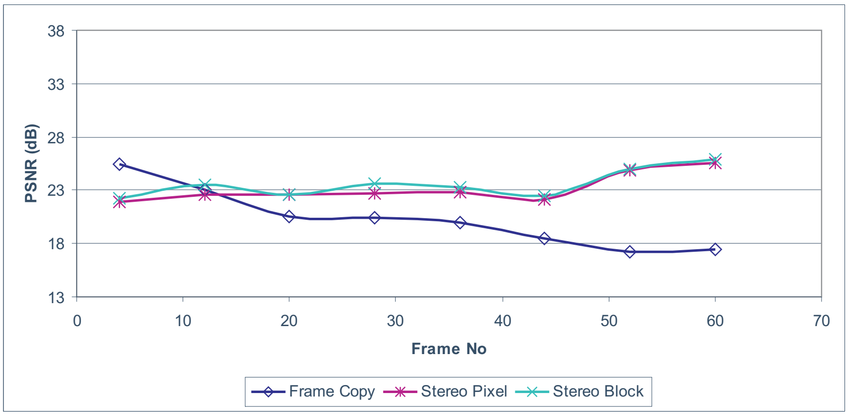


(b) Results of B3 pictures

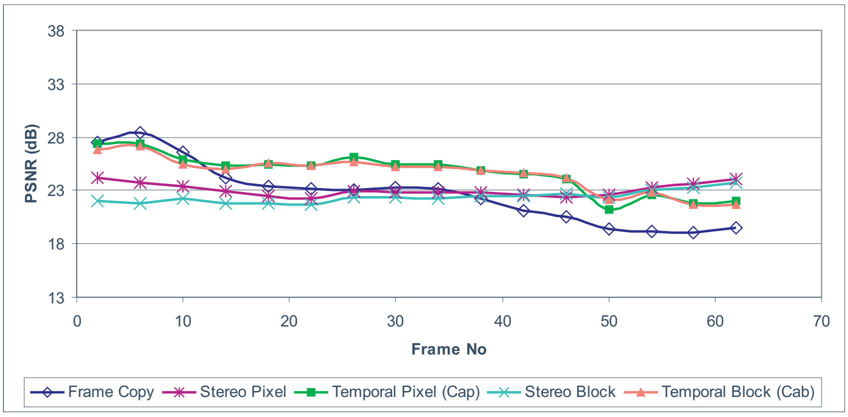


(c) Results of B4 pictures

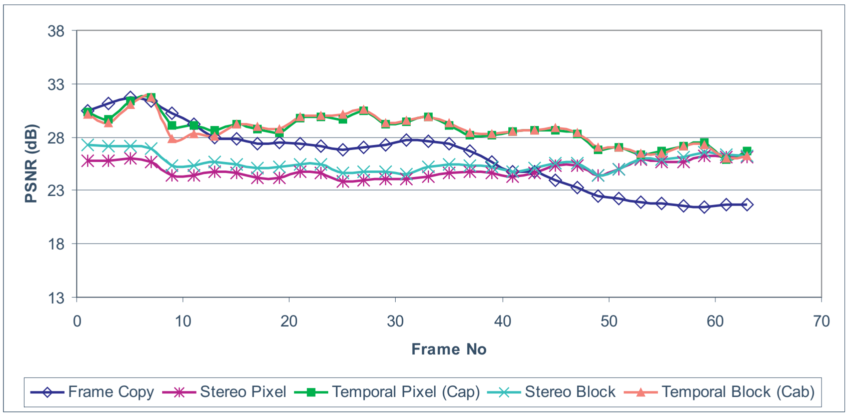
Figure 5.2: Concealment results of Botanical sequence when encoded with Simplified referencing structure of H.264 Multiview Extension (Stereo Pixel = SHBpix, Stereo Block = SHBblk)



(a) Results of B2 pictures

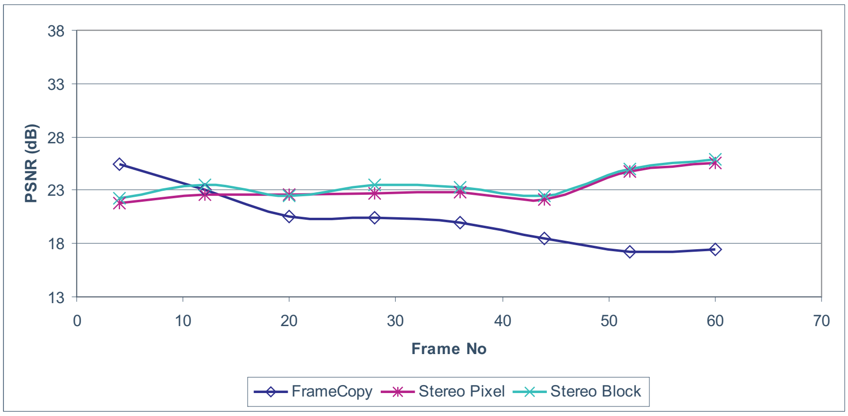


(b) Results of B3 pictures

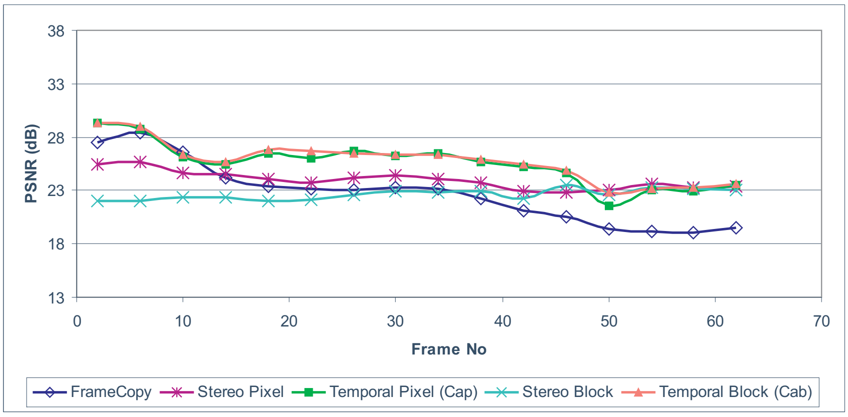


(c) Results of B4 pictures

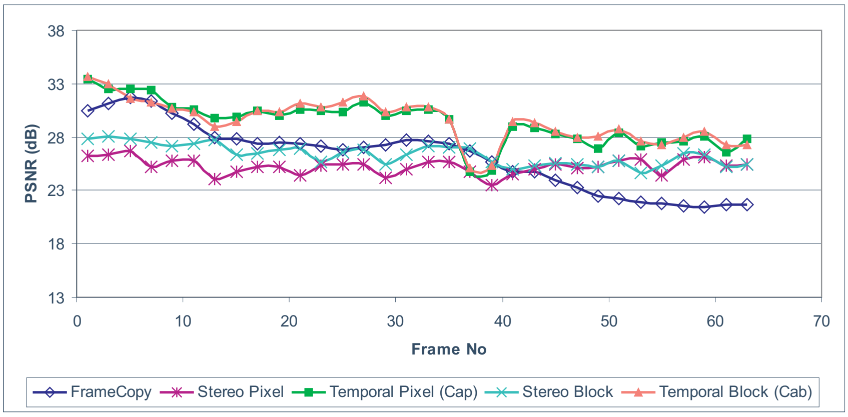
Figure 5.3: Concealment results of Balloons sequence when encoded with General referencing structure of H.264 Multiview Extension (Stereo Pixel = SHBpix, Stereo Block = SHBblk)



(a) Results of B2 pictures

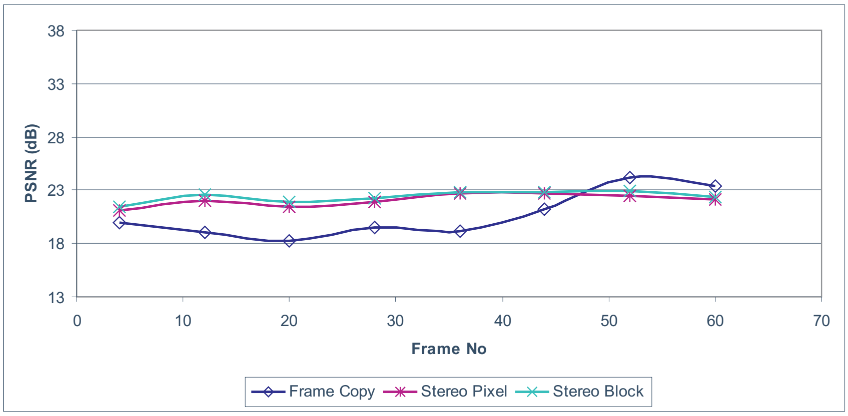


(b) Results of B3 pictures

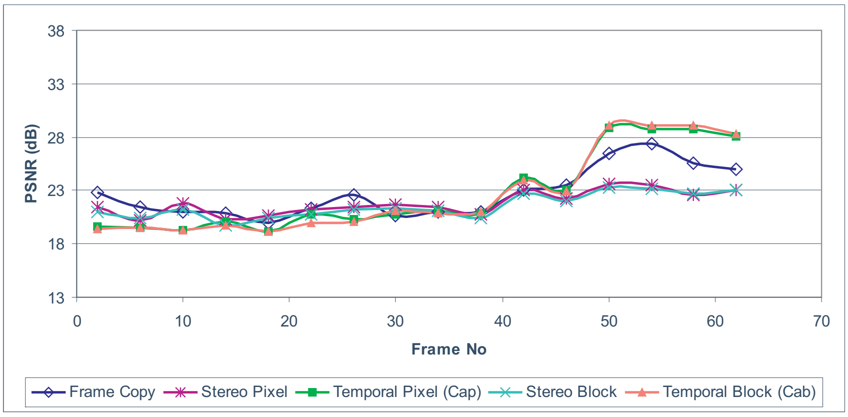


(c) Results of B4 pictures

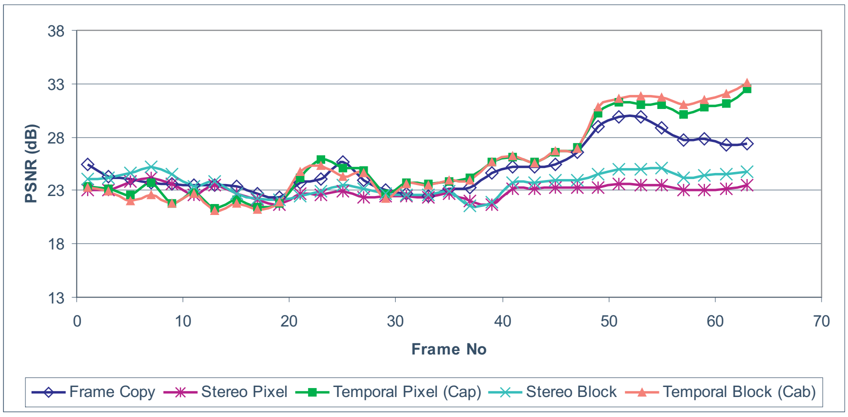
Figure 5.4: Concealment results of Balloons sequence when encoded with Simplified referencing structure of H.264 Multiview Extension (Stereo Pixel = SHBpix, Stereo Block = SHBblk)



(a) Results of B2 pictures

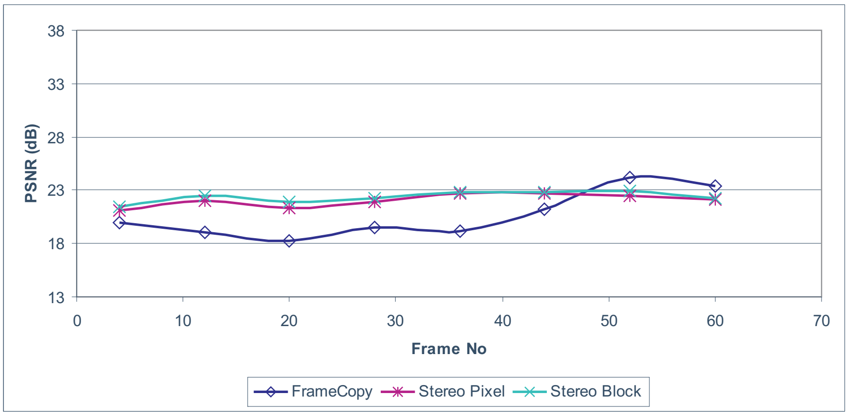


(b) Results of B3 pictures

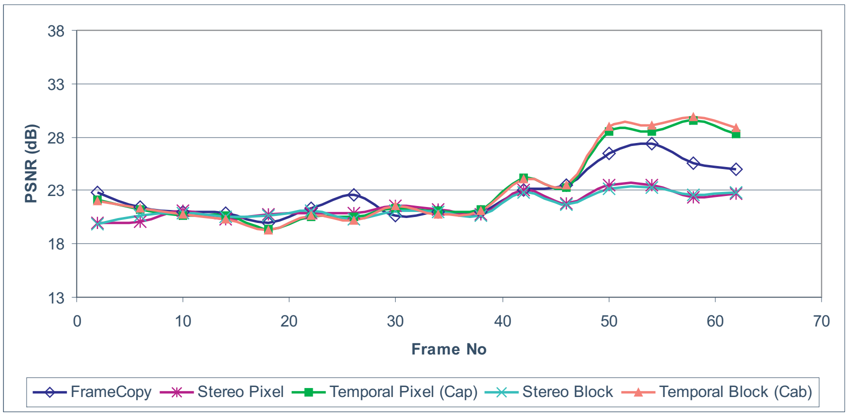


(c) Results of B4 pictures

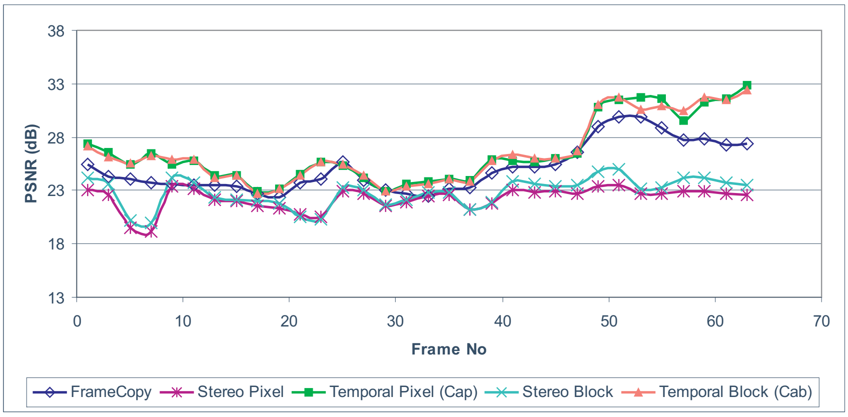
Figure 5.5: Concealment results of Train sequence when encoded with General referencing structure of H.264 Multiview Extension (Stereo Pixel = SHBpix, Stereo Block = SHBblk)



(a) Results of B2 pictures

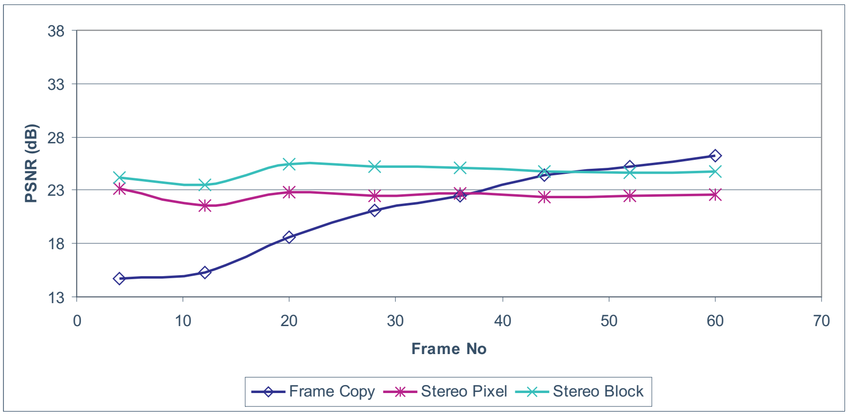


(b) Results of B3 pictures

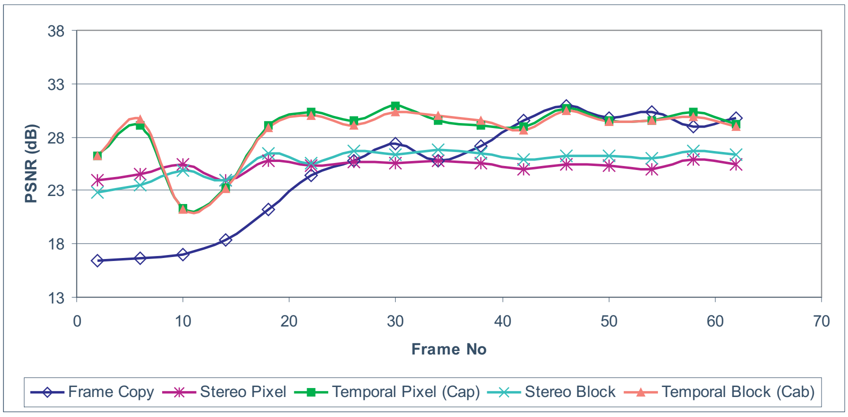


(c) Results of B4 pictures

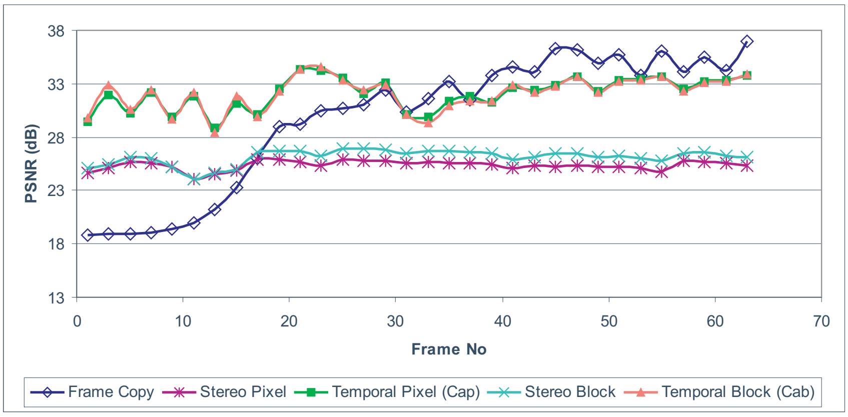
Figure 5.6: Concealment results of Train sequence when encoded with Simplified referencing structure of H.264 Multiview Extension (Stereo Pixel = SHBpix, Stereo Block = SHBblk)



(a) Results of B2 pictures

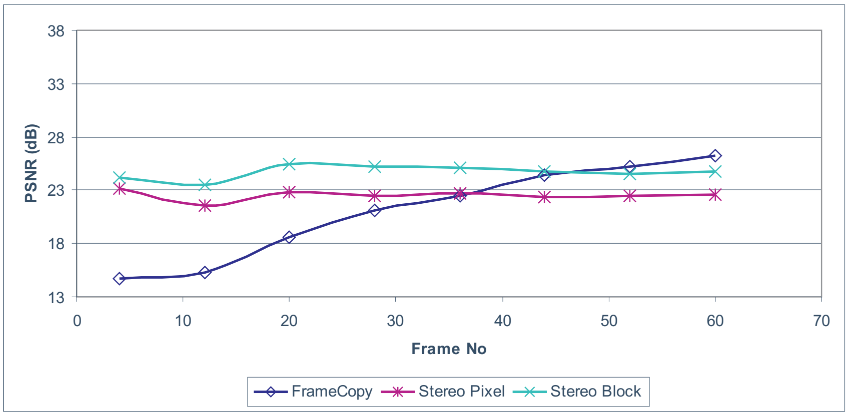


(b) Results of B3 pictures

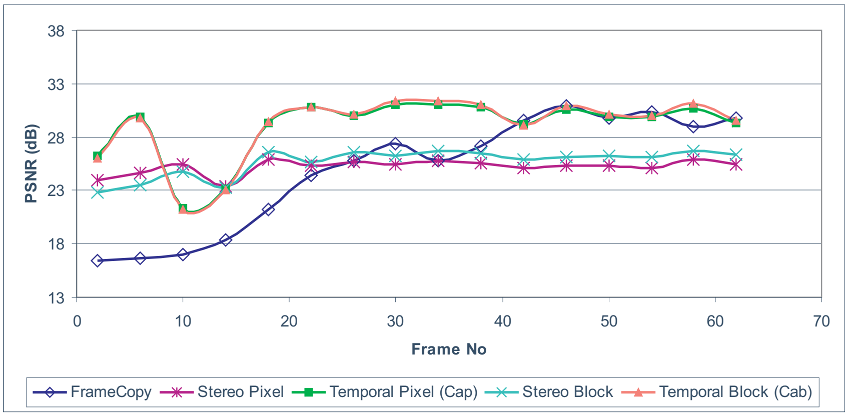


(c) Results of B4 pictures

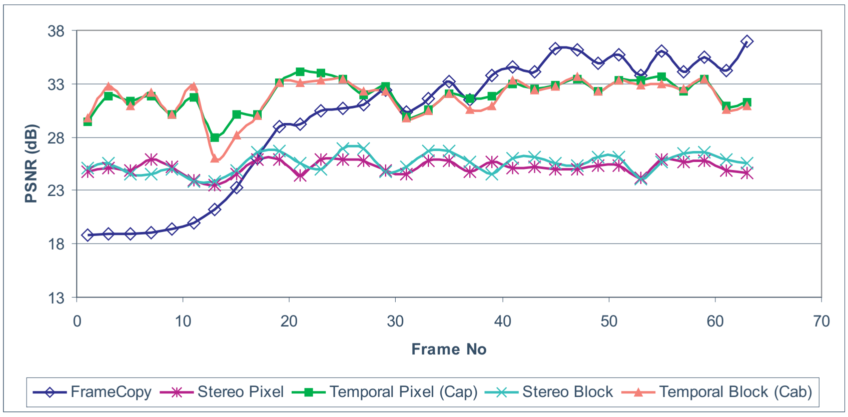
Figure 5.7: Concealment results of Heidelberg sequence when encoded with General referencing structure of H.264 Multiview Extension (Stereo Pixel = SHBpix, Stereo Block = SHBblk)



(a) Results of B2 pictures



(b) Results of B3 pictures



(c) Results of B4 pictures

Figure 5.8: Concealment results of Heidelberg sequence when encoded with Simplified referencing structure of H.264 Multiview Extension (Stereo Pixel = SHBpix, Stereo Block = SHBblk)

5.2 Subjective Quality Comparison Procedure

The subjective quality evaluations are done in order to evaluate the perceptual quality of images or videos. There are several recommendations of ITU for the subjective evaluations and one of these is the Double Stimulus Continuous Quality Scale Test Method ([31]). This method is used to evaluate the subjective algorithm performances in this work because it is recommended for the evaluation of stereoscopic video and images in [31].

5.2.1 Double Stimulus Continuous Quality Scale Test Methodology

In this test method at least 15 assessors, who are chosen among non-experts and inexperienced assessors, should be used. The evaluation should be on a continuous scale ranging from 0 to 100.

The method can be applied in two variants:

Variant1: Each assessor is let to switch between two conditions, A and B (two stereoscopic videos), one of which is always the source and the other is the tested condition applied on the source. The identity of the images, whether it is the source or the test condition, should be known by the experimenter but not by the assessors. After evaluating the conditions the assessor moves to the next pair of images or videos.

Variant2: Multiple assessors are shown two conditions, A and B (two stereoscopic videos), consecutively one of which is always the source and the other is the tested condition applied on the source. The identity of the images, whether it is the source or the test condition, should be known by the experimenter but not by the assessors. The next pair of conditions is shown after the assessors establish an opinion.

Analysis Method: For the analysis of the test results, each evaluation is graded between 0-100 and the difference between the scores of source image and the test condition is calculated to find the score of that test condition on that image by the assessor. After all these scores are calculated, the values are normalized to fit in

0-100. And as a final step, to find the scores of each algorithm (test condition) the average of all the scores over the assessors and images are taken. Scores of the algorithms can be compared with their closeness to the number to which zero score is mapped during the normalization process.

A more detailed explanation of the test method can be found in [31].

5.2.2 Comparisons of Methods on MMRG Multiview Codec with DSCQS Test

For the assessment of the concealment methods on MMRGMC, the videos Soccer, Heidelberg, Akkokayo, Train and Balloons are encoded with MMRGMC with number of cameras equal to two, and also with standard H.264 with the same configurations explained in 5.1.1. And then 21 frames out of 72 frames (the frames 3, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, 35, 38, 41, 44, 47, 50, 53, 56, 59 and 62) are removed from the streams of right view and concealed via the methods Frame Copy, SPpix, CAp, SPblk and CAb methods. For Frame Copy algorithm previous frame of the lost frame is repeated. During the concealment, error propagation is not allowed and the frames other than the lost ones are reconstructed perfectly. In addition to the tested algorithms the original video is also shown to the assessors as a tested video to compare the performance.

For the test procedure a 19" Philips 109S Monitor along with stereoscopic glasses provided by eDimensional Corporation is used. The assessors are informed about the method of assessment, the types of quality factors likely to occur, the grading scale, the sequence and timing conditions of the test. The viewing conditions are arranged according to the recommended settings in [31].

The DSCQS test is applied with variant 2 explained in 5.2.1. The video pairs are shown to the assessors twice in a consecutive order. The test duration was about 15 to 20 minutes for each assessor. 15 assessors (13 male, 2 female with average age 25) with ages ranging from 22 to 31, volunteered to participate in the experiment. The participants were non-experts in the area of picture quality and were screened for

Table 5.14: The mean opinion scores of the algorithms on MMRG Multiview Codec from the DSCQS Test with best quality score of 41

Algorithm	Mean Opinion Score
Original	44.4133
SPpix	59.9867
CAp	63.0667
SPblk	64.2933
CAb	69.1333
Frame Copy	44.7467

color vision, stereo depth perception and visual acuity. At the beginning of the test, 5 random evaluation pairs are shown to the assessors and these 5 evaluation pairs are not evaluated since they provide stabilization of the perception of assessors. The test material is shown in a random order for each assessor. The randomization is done both among evaluation pairs and among the set of video sequences in the pair.

After all the assessors finish the test, the scores are evaluated and normalized. The mean opinion scores (MOS) for each algorithm is shown in Table 5.14. Due to the normalization, 0 (best quality) is mapped to 41, and the success of the algorithms can be measured by closeness of their mean to 41. As it can be seen form the Table 5.14, the score of the original video is not 41 but 44.4133 which is due to misadjustment of the assessors and is expected.

The results in Table 5.14 show that the performance of the SPpix algorithm is superior to both CAp and CAb algorithms and performance of the SPblk algorithm is very close to the CAp algorithm. The score of the Frame Copy algorithm is very close to the score of the original video which means the single frame loss can be tolerated by the assessors. But in case of consecutive frame losses the irritation would increase since the frame gap would be greater (as it will be shown in 5.2.3), also the error propagation would be higher since PSNR of the Frame Copy algorithm is much smaller than other algorithms in most cases.

5.2.3 Comparisons of Methods on H.264 Multiview Extension with DSCQS Test

The methods on H.264 Multiview Extension are tested with the same videos and the same test setup as in 5.2.2. The videos are encoded with both general and simplified schemes of the H.264 Multiview Extension and 19 frames of of 72 (2 (B3), 4 (B2), 10(B3), 13(B4), 15(B4), 18 (B3), 20 (B2), 26(B3), 29(B4), 31(B4), 34 (B3), 36 (B2), 42(B3), 45(B4), 47(B4), 50 (B3), 52 (B2), 58(B3), 61(B4)) are removed from the streams of right view. The SHBpix and SHBblk methods are applied on the general scheme of H.264 Multiview Extension whereas CAp and CAb methods are applied on the simplified scheme. The Frame Copy algorithm is applied by copying the temporally closest available frame to the lost frame at the time of loss (the lost frames are concealed by copying the frames 0, 0, 8, 12, 14, 16, 16, 24, 28, 30, 32, 32, 40, 44, 46, 48, 48, 56, 60 respectively). In addition to the tested algorithms the original video is also shown to the assessors as a tested video to compare the performance.

Table 5.15: The mean opinion scores of the algorithms on H.264 Multiview Extension from the DSCQS Test with best quality score of 43

Algorithm	Mean Opinion Score
Original	49.68
SPpix	61.0133
CAp	60.6
SPblk	61.5467
CAb	61.8667
Frame Copy	62.4933

15 assessors (12 male, 3 female with average age 24) with ages ranging from 20 to 31, volunteered to participate in the experiment. The participants were non-

experts in the area of picture quality and were screened for color vision, stereo depth perception and visual acuity. After all the assessors finish the test, the scores are evaluated and normalized. The mean opinion scores (MOS) for each algorithm is shown in Table 5.15. Due to the normalization, 0 (best quality) is mapped to 43, and the success of the algorithms can be measured by closeness of their mean to 43.

The results in Table 5.15 show that the temporal and spatial algorithms does not provide a significant difference in the performance in H.264 Multiview Extension and the comparisons should be made with objective measures. However these results are significant in the sense that the Frame Copy algorithm performance degraded significantly with respect to the ones presented in Table 5.14. This is due to frame gaps between the lost frame and copied frame in case of first and second level non-anchor B frames. Therefore the Frame Copy algorithm in MMRG Multiview codec would also suffer from this effect in case of multiple consecutive frame losses.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

In this thesis, the full frame loss methods on multiview videos are investigated and four methods are proposed to use the inter-view compensations in the multiview codecs. The method SPpix is a full frame loss concealment method that basically uses both temporal compensation and disparity compensation information in the previous frames to efficiently conceal the lost frames. The algorithm outperforms the Frame Copy algorithm and temporal motion vector projection algorithms (CAp and CAb) even if they are applied to the videos that are only temporally compensated. The proposed SPpix algorithm works in pixel domain and its complexity is high for applications of real-time streaming especially with high resolution videos. SPblk algorithm is proposed in order to decrease the complexity of the SPpix algorithm without degrading its performance. Both subjective and objective evaluation results show that SPblk algorithm has a close performance to SPpix algorithm and it can be used in real-time applications.

The methods SHBpix and SHBblk are proposed to be used with H.264 Multiview Extension, a codec that is proposed to be the standard for multiview coding by ITU and ISO. SHBpix algorithm uses the motion vectors of the previously decoded frames to form an estimate disparity field from the disparity vectors of the previous frames and anchor frames. This disparity field is then used to reconstruct the concealed picture. SHBblk algorithm is proposed as the block based version of SHBpix, that works in blocks rather than pixels to reduce the complexity. The methods SHBpix and SHBblk are shown to be have very similar performances. The objective evaluations showed that the performance of the methods change with respect to the type of the picture to be concealed, i.e. the concealment of B pictures at different levels presented different results. For the concealment of first level non-anchor B pictures, the proposed methods outperformed Frame Copy which is the only available method for this type of pictures. For the concealment of second level non-anchor B pictures the results show that the proposed methods are superior to temporal methods unless the video to be concealed does not have a stationary scene with slowly moving objects. In the case of high motion and camera motion the proposed algorithms performed better than temporal methods. The concealment of third level non-anchor B pictures resulted in worst performance of proposed algorithms with respect to temporal algorithms and it is shown that only in general scheme of H.264 Mıltiview Extension, i.e. only when there are inter-view references of each frame, and in high motion videos the proposed algorithms can be used instead of temporal algorithms. The SHBblk algorithm is also shown to be suitable for using in real-time applications.

It can be inferred from the results in this thesis that in case of coding the multiview sequences with P-pictures as in MMRG Multiview Codec, the proposed SPblk and SPpix algorithms can be used for full frame concealment. Another conclusion is that as the number of levels in the hierarchical B-pictures increase, the performance of the SHBpix and SHBblk algorithms will be superior to temporal algorithms in the lower levels. Also an adaptive algorithm can be used to understand the video characteristics and select the suitable algorithm for full frame loss concealment. For instance the percentage of non-zero temporal references in the previous frame can be used to determine the stationarity of the scene and temporal algorithms may be used for stationary scenes. Another improvement can be added for the general scheme of the H.264 Multiview Extension that the SPpix and SPblk algorithms can be applied to the hierarchical B-pictures. This will enable the use of motion vectors to partially reconstruct the concealed picture and will improve the results in the stationary scenes. However these methods would not provide any improvement over temporal methods when the percentage of inter-view compensated blocks are low. They are also not applicable in case of simplified scheme.

The multiview codecs are evolving in order to encode the pre-generated depth data in addition to the video. In such a codec structure where the original depth information is available for each frame of the encoded video, the proposed methods in this thesis can be used with original depth instead of inter-view references which may provide better performances compared to codec structures without the depth information.

REFERENCES

[1] R.M. Mersereau Y.-C. Lee, Y. Altunbasak, “Multiframe error concealment for mpeg-coded video delivery over error-prone networks,” November 2002, vol. 11 of *Image Processing, IEEE Transactions on*, pp. 1314 – 1331.

[2] L. Shaw Y. Wang, Q.-F. Zhu, “Maximally smooth image recovery in transform coding,” October 1993, vol. 41 of *Communications, IEEE Transactions on*, pp. 1544 – 1551.

[3] M. Ghanbari T. Shanableh, “Loss concealment using b-pictures motion information,” June 2003, vol. 5 of *Multimedia, IEEE Transactions on*, pp. 257 – 266.

[4] S. N. Merchant S. Garg, “Interpolated candidate motion vectors for boundary matching error concealment technique in video,” October 2006, vol. 53 of *Circuits and Systems II: Express Briefs, IEEE Transactions on*, pp. 1039 – 1043.

[5] Y.-S. Ho J.-W. Suh, “Error concealment techniques for digital tv,” December 2002, vol. 48 of *Broadcasting, IEEE Transactions on*, pp. 299 – 306.

[6] C.N. Canagarajah D. Agrafiotis, D.R. Bull, “Enhanced error concealment with mode selection,” August 2006, vol. 16 of *Circuits and Systems for Video Technology, IEEE Transactions on*, pp. 960 – 973.

[7] T.W. Yang Q. Peng and C.Q. Zhu, “Block-based temporal error concealment for video packet using motion vector extrapolation,” 2002, Proc. IEEE Communications, Circuits and Systems and West Sino Expositions.

[8] E. Magli G. Olmo S. Belfiore, M. Grangetto, "Concealment of whole-frame losses for wireless low bit-rate video based on multiframe optical flow estimation," April 2005, vol. 7 of *Multimedia, IEEE Transactions on*, pp. 316 – 329.

[9] A. Chimienti L. Pezzoni F.S. Rovati P. Baccichet, D. Bagni, "Frame concealment for h.264/avc decoders," February 2005, vol. 51 of *Consumer Electronics, IEEE Transactions on*, pp. 227 – 233.

[10] M. Kunter T. Sikora S. Knorr, C. Clemens, "Robust concealment for erroneous block bursts in stereoscopic images," 3DPVT 2004, September 2004.

[11] Joint Video Team, "Advanced video coding for generic audiovisual services," Tech. Rep. ITU-T Recommendation H.264 - ISO/IEC 14496-10 AVC, ITU-T and ISO/IEC JTC1, 2003.

[12] G. Jiang T.-Y. Choi Y.-D. Kim S. Li, M. Yu, "Approaches to h.264-based stereoscopic video coding," in *Proceedings of the Third International Conference on Image and Graphics*, December 2004.

[13] C. Bilen A. Aksay, G. B. Akar, "A multi-view video codec based on h.264," 2006 IEEE International Conference on Image Processing, October 2006, pp. 541 – 544.

[14] Joint Video Team, "Joint draft 1.0 on multiview video coding," Tech. Rep. JVT-U209, ITU-T and ISO/IEC JTC1, November 2006.

[15] Joint Video Team, "Joint multiview video model (jmvm) 2.0," Tech. Rep. JVT-U207, ITU-T and ISO/IEC JTC1, October 2006.

[16] Joint Video Team, "Comparative study of mvc prediction structures," Tech. Rep. JVT-V132, ITU-T and ISO/IEC JTC1, January 2007.

[17] T. Wiegand P. Merkle A. Smolic, K. Müller, "Coding efficiency and complexity analysis of mvc prediction structures," EUSIPCO, 2007.

[18] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: A review," May 1998, vol. 86 of *Proceedings of the IEEE*, pp. 974 – 997.

[19] Y.-L. Huang and H.-Y. Lien, "Temporal error concealment for mpeg coded video using a self-organizing map," May 2006, vol. 52 of *Consumer Electronics, IEEE Transactions on*, pp. 676 – 681.

[20] C.-S. Chen M.-J. Chen and M.-C. Chi, "Temporal error concealment algorithm by recursive block-matching principle," November 2005, vol. 15 of *Circuits and Systems for Video Technology, IEEE Transactions on*, pp. 1385 – 1393.

[21] C.-C.J. Kuo W.-Y. Kung, C.-S. Kim, "Spatial and temporal error concealment techniques for video transmission over noisy channels," July 2006, vol. 16 of *Circuits and Systems for Video Technology, IEEE Transactions on*, pp. 789 – 803.

[22] Y. Zhou Y. Xu, "H.264 video communication based refined error concealment schemes," November 2004, vol. 50 of *Consumer Electronics, IEEE Transactions on*, pp. 1135 – 1141.

[23] S.-W. Chou S.-C. Hsia, S.-C. Cheng, "Efficient adaptive error concealment technique for video decoding system," October 2005, vol. 7 of *Multimedia, IEEE Transactions on*, pp. 860 – 868.

[24] P.C. Cosman S. Chen, "Decision trees for error concealment in video decoding," March 2003, vol. 5 of *Multimedia, IEEE Transactions on*, pp. 1 – 7.

[25] J.M. Boyce Z. Wu, "An error concealment scheme for entire frame losses based on h.264/avc," ISCAS 2006, May 2006.

[26] J. Li S. Li Y. Chen, K. Yu, "An error concealment algorithm for entire frame loss in video transmission," Picture Coding Symposium 2004, December 2004.

[27] C. Clemens T. Sikora M. Kunter, S. Knorr, "A gradient based approach for stereoscopic error concealment," ICIP 2004, October 2004.

[28] C. Clemens T. Sikora M. Kunter, S. Knorr, "A hybrid approach for error concealment in stereoscopic images," WIAMIS 2004, 2004.

[29] G. Jiang Z. Jiang F. Li L. Pang, M. Yu, "An approach to error concealment for entire right frame loss in stereoscopic video transmission," Computational Intelligence and Security, 2006 International Conference on, November 2006.

[30] Jiang G. Choi T. Li S., Yu M., "Approaches to h.264 based stereoscopic video coding," Proceedings of Int. Conf. on Image and Graphic, December 2004.

[31] "Methodology for the subjective assessment of the quality of television pictures," Tech. Rep. Recommendation BT.500-10, ITU, 2000.