ERROR CONCEALMENT IN 3D VIDEO

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

BY

SERCAN AYDOĞMUŞ

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

DECEMBER 2011

Approval of the thesis

ERROR CONCEALMENT IN 3D VIDEO

Submitted by **Sercan AYDOĞMUŞ** in partial fulfillment of the requirements for the degree of **Master of Science in Electrical and Electronics Engineering Department, Middle East Technical University** 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 Dept., METU**

Examining Committee Members:

Prof. Dr. Aydın Alatan ____________
Electrical and Electronics Engineering Dept., METU

Prof. Dr. Gözde Bozdağı Akar ____________
Electrical and Electronics Engineering Dept., METU

Assoc. Prof. Dr. Mehmet Mete Bulut ____________
Electrical and Electronics Engineering Dept., METU

Assoc. Prof. Dr. Çağatay Candan ____________
Electrical and Electronics Engineering Dept., METU

Barış Konuk, M.Sc. ____________
REHİS, ASELSAN

Date: 8.12.2011

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 : Sercan Aydoğmuş

Signature :

ABSTRACT

ERROR CONCEALMENT IN 3D VIDEO

Aydoğmuş, Sercan

M. Sc., Department of Electrical and Electronics Engineering

Supervisor: Prof. Dr. Gözde Bozdağı Akar

December 2011, 96 pages

The advances in multimedia technologies increased the interest in utilizing three dimensional (3D) video applications in mobile devices. However, wireless transmission is significantly prone to errors. Typically, packets may be corrupted or lost due to transmission errors, causing blocking artifacts. Furthermore, because of compression and coding, the error propagates through the sequence and salient features of the video cannot be recovered until a key-frame or synchronization-frame is correctly received. Without the use of concealment and enhancement techniques, visible artifacts would inevitably and regularly appear in the decoded stream. In this thesis, error concealment techniques for full frame losses in depth plus video and stereo video structures are implemented and compared. Temporal and interview correlations are utilized to predict the lost frames while considering the memory usage and computational complexity.

The concealment methods are implemented on jm17.2 decoder which is based on H.264/AVC specifications [1]. The simulation results are compared with the simple frame copy (FC) method for different sequences having different characteristics.

Keywords: Error concealment, full frame loss, depth plus video, stereo video, H.264/AVC.

ÖZ

3B VİDEODA HATA ÖRTME

Aydoğmuş, Sercan

Yüksek Lisans, Elektrik Elektronik Mühendisliği Bölümü

Tez Yöneticisi: Prof. Dr. Gözde Bozdağı Akar

Aralık 2011, 96 sayfa

Multimedya teknolojilerindeki gelişmeler üç boyutlu (3B) video uygulamalarının taşınabilir cihazlarda kullanımı üzerindeki ilgiyi artırdı. Ancak, kablosuz iletim hataya önemli derecede yatkındır. Tipik olarak paketler iletim hatalarından dolayı bozulabilir ya da kaybolabilir ve yapay blok etkisine neden olur. Ayrıca, sıkıştırma ve kodlamadan dolayı, hata seri boyunca yayılır ve bir anahtar-film karesi ya da eşleme-film karesi hatasız olarak alınmadığı takdirde videodaki belirgin hatlar düzeltilemez. Örtme ve geliştirme teknikleri kullanılmadan, açık yapaylıklar kaçınılmaz ve düzenli bir şekilde çözülmüş film karesinde görünür. Bu tezde, derinlik artı video ve çift-görüşlü video yapılarında tam film karesi kayıpları için hata örtme teknikleri uygulanmakta ve karşılaştırılmaktadır. Zamansal ve görüş arası bağıntılar, hafıza kullanımı ve hesaplama karmaşıklığı da göz önünde bulundurulurulurken, kayıp tam film karesi kayıplarını tahmin etmek için kullanılmıştır.

Hata örtme yöntemleri H.264/AVC tanımlaması [1] tabanlı JM.17.2 çözücüsünde uygulanmıştır. Simülasyon sonuçları basit film karesi kopyalama yöntemi ile karşılaştırılmıştır.

Anahtar kelimeler: Hata örtme, tüm film karesi kaybı, derinlik artı video, çift-görüşlü video, H.264/AVC.

To My Family

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest gratitude to my supervisor Prof. Dr. Gözde Bozdağı Akar for her sincerity, encouragement and support throughout my graduate study.

I would like to thank ASELSAN Inc. for facilities provided for the completion of this thesis.

I would also like to thank TÜBİTAK-BİDEB for their financial support during my graduate study.

I would like to express my deepest thanks to my family for their continuous support and encouragements.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

2D	: 2 Dimensional
3D	: 3 Dimensional
ARQ	: Automatic Retransmission Query
AVC	: Advanced Video Coding
CMVC	: Complementary Motion Vector Copy
CMVDVC	: Combined Motion Vector and Disparity Vector Copy
DCT	: Discrete Cosine Transform
DV	: Disparity Vector
DVC	: Disparity Vector Copy
EC	: Error Concealment
FC	: Frame Copy
FCBR	: Frame Copy Before Rendering
FEC	: Forward Error Correction
GOP	: Group of Pictures
MAD	: Mean Absolute Difference
MB	: Macro Block
MPEG	: Moving Picture Experts Group
MV	: Motion Vector
MVC	: Multiview Video Coding
NAL	: Network Abstraction Layer
NALU	: Network Abstraction Layer Unit

POCS	: Projections On to Convex Sets
PSNR	: Peak Signal to Noise Ratio
RANSAC	: Random Sampling Consensus
SSIM	: Structural Similarity
VCL	: Video Coding Layer

CHAPTER 1

INTRODUCTION

1.1 GENERAL

The evolution of the video coding standards made it possible to adapt multimedia technologies to hand-held devices. Video streaming, video conferencing, video publishing are some of the applications which can be utilized in mobile phones, smart phones, personal assistants and so on.

The introduction of the latest coding standard and the achievements in three dimensional (3D) display systems renewed the interest on the 3D video. As more products become available for the consumer markets, 3D video for mobile devices gain popularity such as mobile 3D-TV, 3D video conferencing with cell phones and so on.

3D vision, or 3D video phrases are commonly used for the visual media providing depth perception to the viewer. The simplest form of the 3D video is stereoscopic video. Display subsystem of a stereo video system provides 3D perception by providing special different views to each eye of the observer. The capture subsystem of a stereo video system consists of two camera placed apart. The distance between the two cameras is mostly close to the average distance between two eyes of a human being. A sample stereo system is given in Figure 1-1. Multiview video, being a superset of stereo video, is obtained by extending the capture and display subsystems i.e. the number of cameras is increased and the display subsystem can provide more than two different angles of a scene.

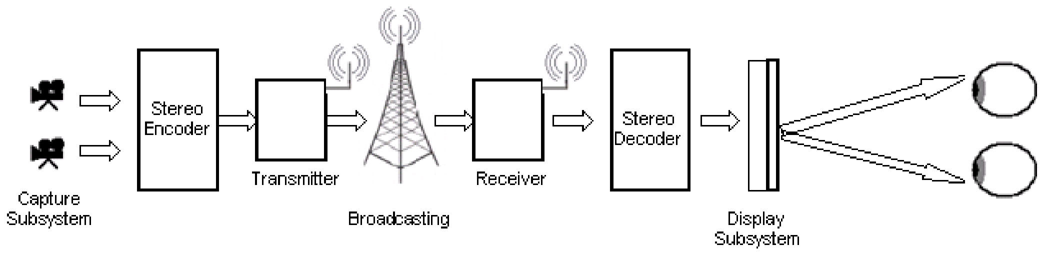


Figure 1-1: Sample stereo video system

For different representations of 3D video, such as stereoscopic, multiview, multiview plus depth, layered depth and depth plus video, different coding techniques are needed. Coding with state of the art monoscopic techniques would increase the bit-rates. Therefore multiview video coding (MVC), has been a popular research area and the latest monoscopic coding standard, H.264/AVC, is extended to support MVC structures [1].

In video communications, error prone channels are very likely to cause packet losses. Since for low bit-rate applications a packet usually contains a large portion or the whole frame, the packet losses need to be recovered efficiently. Forward error correction (FEC) or automatic retransmission query (ARQ) and error concealment (EC) techniques are the candidates for the solution. However FEC and ARQ are not practical because FEC increases the bitrates and ARQ increases the delays. As for EC, there is no need for any error correction headers, or retransmission. EC techniques do not require extra information, and predict the lost frames using the previous/next frames or the successfully received surrounding blocks.

There are quite few EC algorithms proposed for monoscopic video decoders in the literature. However those cannot be directly applied to the 3D case. For different representations, appropriate concealment techniques are necessary.

1.2 SCOPE OF THESIS

In this thesis, error concealment methods have been implemented to recover full frame losses. The methods are implemented by modifying the source code of jm17.2 H264/AVC decoder software. The aim is to provide a solution which does not increase memory or computational power needs so that the algorithm can run on mobile platforms too.

Separate concealment algorithms are implemented for depth plus video and stereo video representations. The results are compared with the low-cost Frame Copy (FC) method. The results are given using the objective metrics Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM). They are also evaluated subjectively.

For the depth plus video sequences, two different concealment methods are implemented. The first method is frame copy before rendering (FCBR), and the second one is complementary motion vector copy (CMVC) method.

For the stereo video sequences, three concealment algorithms are implemented. The first method is complementary motion vector copy (CMVC), the second one is disparity vector copy (DVC) and the third is combined motion vector and disparity vector copy (CMVDVC) method.

1.3 THESIS OUTLINE

The thesis is organized as follows. In Chapter 2, basics of the 2D+depth multiview video coding and coding are introduced.

In Chapter 3, literature review of error concealment for monoscopic video sequences and stereoscopic video sequences are given.

The implemented algorithms for depth plus video and stereo video representations are explained in Chapter 4.

In Chapter 5, the graphical results and examples for the concealed frames are presented. The results are compared and the performances are evaluated.

Finally, in Chapter 6, the conclusions and future improvements are given.

CHAPTER 2

3D VIDEO CODING

2.1 GENERAL

3D video can be represented in different formats. Teo of the most commonly used formats are multiview representation and 2D+depth representation. 2D+depth representation consists of the conventional color video sequence and corresponding depth sequence. The depth images are the 2D grey scale representations for the 3D surface of the scene, where the pixels takes the values according to the distance from the camera. The sequences of the depth images need to be coded efficiently similar to the color video sequences. On the otherhand multiview video is the video format that contains multiple video sequences captured simultaneously by several cameras placed at different locations. Stereo video is a special case of multiview video, where the number of cameras is two. Due to the number of the cameras, multiview video contains large amount of raw bit-rate of video. Therefore efficient compression techniques are required for 3D video communications systems. In this chapter, basic principles of multiview video coding and 2D+depth coding that affects the concealment techniques to be proposed are given.

2.2 2D+DEPTH CODING

2D+depth representation is the format where a video sequence is coded with the corresponding depth map sequence. The 2D+depth video representation is used in 3D systems to synthesize different angle of views for the scene.

The depth range is restricted to a minimum and maximum distance from the camera. Since the depth map corresponds to a monochromatic image, the sequence of those maps can be converted to a 4:0:0 YUV format video sequence.

The 2D+depth sequences are coded with the state of the art monoscopic video coding standards. The simulcast coding structure is utilized where the two sequences are coded as separate monoscopic video sequences.

2.3 MULTIVIEW VIDEO CODING

With the introduction of H.264 Advanced Video Coding (AVC) standard, significant improvements have been achieved in monoscopic video coding [2]. H.264 Advanced Video Coding standard is fundamental for multiview video coding. However, when the strong interview relevancy and the amount of data of multiview video are considered, monoscopic coding standard becomes insufficient. Therefore, in 2008, multiview extension for the H.264 is proposed as H.264/Multiview Video Coding (MVC) where interview redundancy is considered in compression [3]. The multiview video coding is standardized as the extension of H.264/MPEG-4 AVC in 2010 [4].

2.3.1 H.264/MPEG-4 AVC

The main building blocks of the MVC encoder/decoder are similar to the ones given in the state of the art H.264 monoscopic video coding standard. The two basic blocks of H.264 encoder/decoder are Network Abstraction Layer (NAL) and Video Coding Layer (VCL).

In the Network Abstraction Layer, the coded video data and/or the additional informations about the video stream are prepared as NAL units (NALU). Each NALU can contain either some portion (slice) or even a full coded frame. Therefore erroneously received or lost NALUs may cause full frame losses which need to be handled in concealment phase of decoder.

In the VCL, the frames are partitioned as smaller coding units named as Macroblocks (MB). Each MB is predicted spatially or temporally according to the slice type. There are three types of slices. The first one is the I-slice, where each MB of the I-slice uses intrapicture coding. In intrapicture coding, MBs are spatially predicted from the neighbouring MBs in the current slice. Therefore I-slices are self decodable where, no other video data from other slices/frames are needed. The second one is P-slice, where each MB of the P-slice uses either intrapicture coding or interpicture predictive coding. In interpicture predictive coding, MBs are predicted temporally from one of the previously decoded frames. The third one is B-slice, where each MB of the slice uses either intrapicture coding or interpicture predictive coding and interpicture bipredictive coding. In interpicture bipredictive coding, MBs are predicted temporally from previous and future decoded frames. The two predictions from the previous and future frames are combined with a weighted average. Figure 2-1 shows the prediction schemes.

The Hierarchial B Pictures provides flexibility to the coding order, i.e. the display order and the decode order can be different. The reference for a slice can be any of the frames that are decoded previously and stored in the Decoded Picture Buffer (DPB). Slices are labeled with temporal levels i.e. I and P slices are level 0 pictures and B slices are higher level pictures due to the importance of I and P slices. Level 0 is the most important layer. Each B slice is predicted from the closest lowest level pictures. The prediction structure repeats itself within a number of pictures. This repetition is named as Group of Pictures (GOP). Figure 2-2 shows an example of hierarchial B structure having 8 pictures in a GOP. In the figure, the coding order is: 1, 9, 5, 3, 7, 2, 4, 6, 8 and so on.

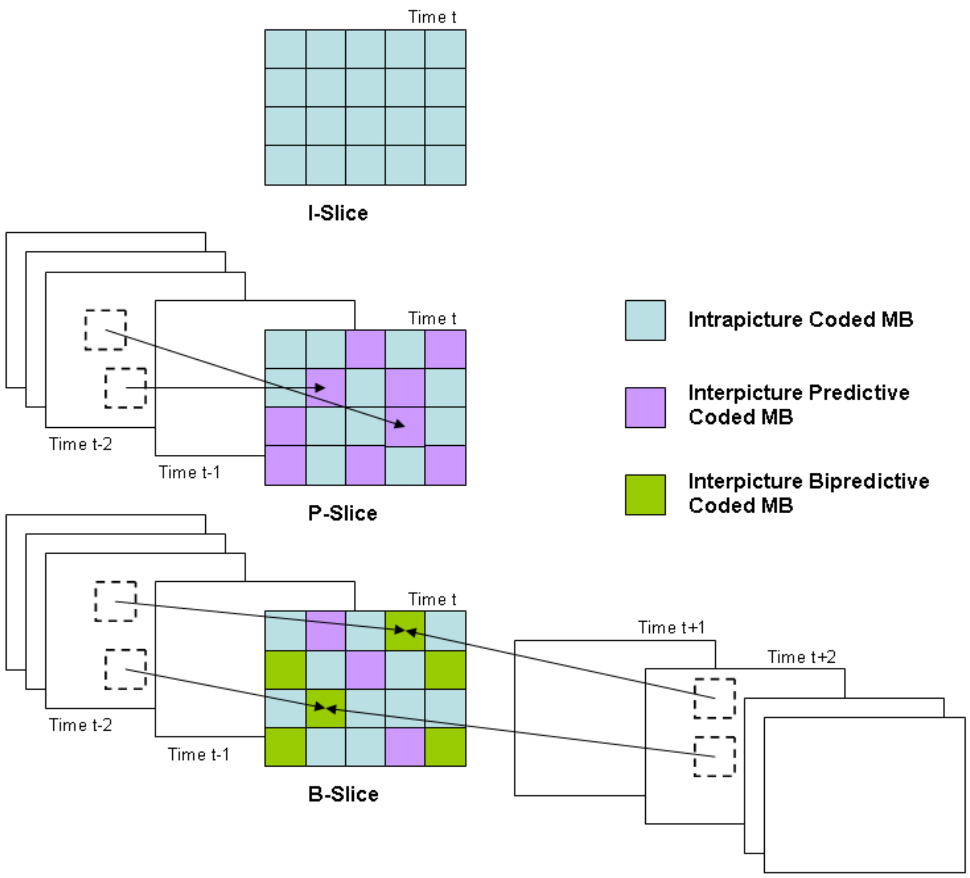


Figure 2-1: H.264 prediction schemes

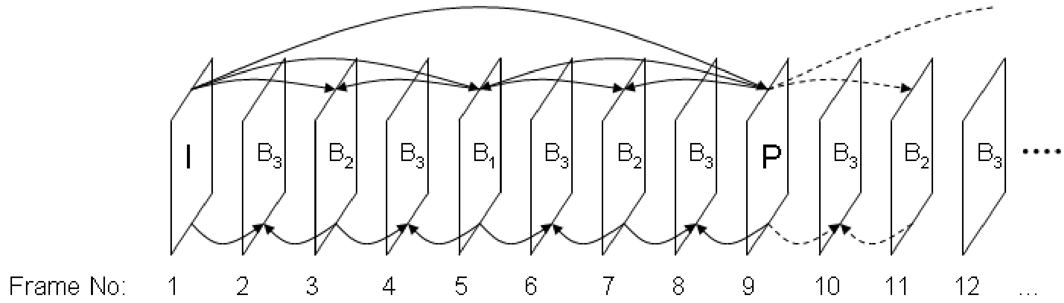


Figure 2-2: Hierarchial B picture coding structure for number (GOP) = 8

In H.264/MPEG-4 AVC, the MBs are further divided into smaller coding portions, named as blocks. The intra coded 16x16 luma MBs can be divided into 4x4 or 8x8 blocks for individual intrapicture predictions. The inter coded (predictive and bipredictive) 16x16 luma MBs can be divided into 16x8, 8x16, 8x8, 8x4, 4x8, and 4x4 blocks for individual interpicture predictions.

2.3.2 MVC EXTENSION

In MVC scheme, one of the views is chosen to be the base view. The base view is encoded as a mono video sequence, which coded with AVC standard, and organized as separate NALUs. The separation of the base view and other views provides backward compatibility i.e. a 2D system can filter the base view's packets and decode 2D video stream.

In addition to the temporal and spatial correlations, there are strong correlations between the views. For the high compression, in MVC, inter view prediction is utilized. The displacement vectors that are generated from the inter view prediction are named as Disparity Vectors (DV). Since H.264 allows flexible referencing with Hierarchical B Pictures, disparity estimation/compensation is not different from motion estimation/compensation algorithmically.

The most straightforward coding structure for the multiview video is to code the individual views as mono sequences. This structure is called as Simulcast Coding and no inter view prediction is involved. The scheme is shown in the Figure 2-3. Another coding structure is given in Figure 2-4. In this structure, interview prediction is utilized for the key pictures. The remaining frames are coded similar to the mono case. In the Figure 2-5, the hierarchial B prediction structure is used for all frames. The coding complexity is high for the last prediction structure and memory requirement increases due to the number of reference frames that are needed to be preserved in the decoded picture buffer.

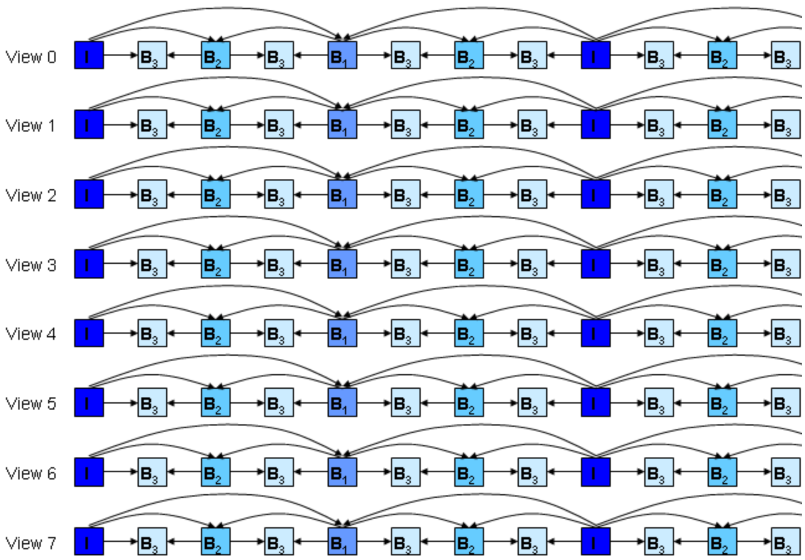


Figure 2-3: Simulcast (Temporal) Coding

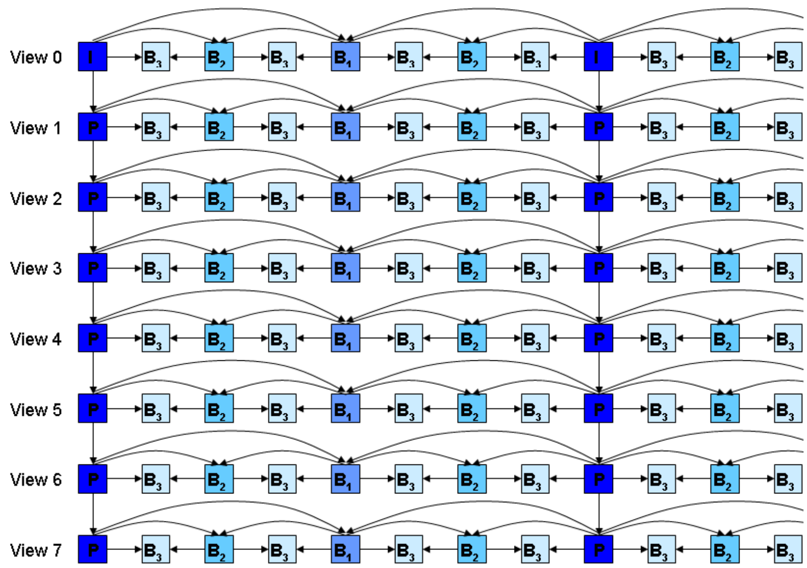


Figure 2-4: Inter View Prediction For Key Pictures

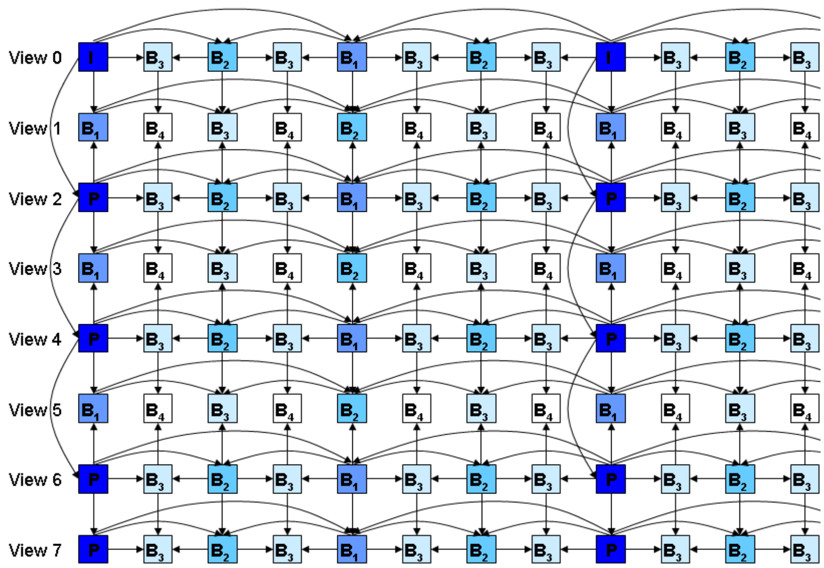


Figure 2-5: Inter View Prediction For All Pictures

Stereo video is a special case of multiview video where the number of the views is 2. The coding structures given for the multiview video are applied to the stereo video too. In this thesis, the proposed error concealment methods are implemented for stereo video coding structures. For both stereo and 2D+depth cases, jm-17.2 H.264/AVC decoder is used to implement error concealment algorithms.

CHAPTER 3

ERROR CONCEALMENT ALGORITHMS

3.1 INTRODUCTION

Video transmission over a medium often results in packet losses caused by bit errors, path losses, congestion, fading effects and so on. Even a single bit error in a packet causes the whole slice or frame to be lost because of strong compression techniques used in the systems. There are several approaches to overcome the packet losses such as automatic retransmission on request [5], layered coding [6], forward error correction (FEC) [7] and error concealment (EC). Almost all of the approaches handle the errors by increasing the redundancy which also increases the transmitted bit stream size. Some of them are not applicable to all systems because of the need of modification in the encoder, or introduction of transmission delay which are not preferred in real time video applications. However, EC techniques do not require any modifications in the encoder side, and does not cause redundancy or transmission delays in the bit stream. Error concealment is simply filling a lost area in the video by predicting it from already received data Therefore EC can recover the errors which cannot be corrected by FEC techniques and does not require additional information, and redundancy in the bit stream. Mostly the only modification to be made is on the decoder side. The data that is not received correctly cannot be recovered perfectly. However human eye can tolerate some degree of error and EC techniques performs well.

Depending on the information used for concealment, EC techniques can be categorized as Spatial, Temporal and Hybrid. The spatial correlation techniques use the high spatial correlation, whereas the temporal concealment techniques use the temporal correlations. The hybrid concealment techniques are the combinations of spatial and temporal techniques. For stereo video, in addition to spatial and temporal correlations, between the views there exist interview correlations. Those correlations also contribute to improve the error concealment algorithms along with the old fashioned monoscopic concealment techniques.

Although over 20 years error concealment techniques are the matter of investigations, even the studies on error concealment for mono video systems are not still complete. Those studies are the foundations of concealment techniques for the stereo video systems. In the following sections, first monoscopic and then stereoscopic concealment techniques are to be reviewed.

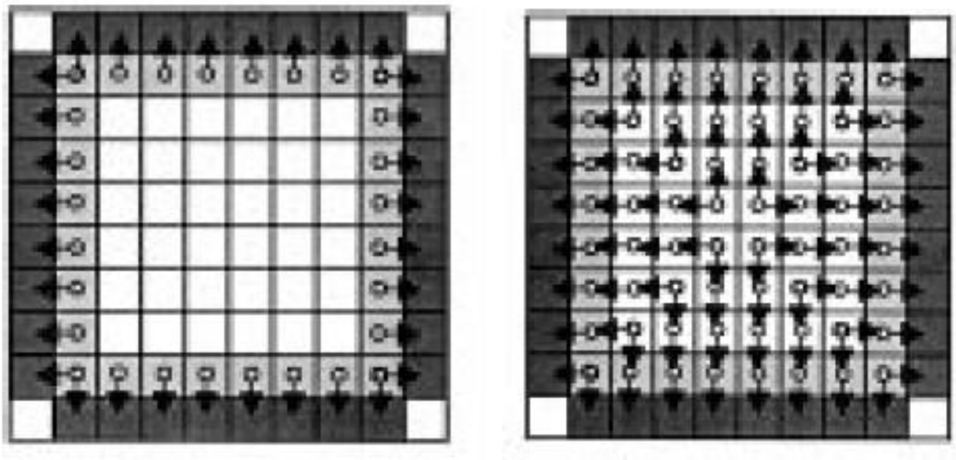
3.2 ERROR CONCEALMENT FOR MONOSCOPIC VIDEO

In a video stream, when a block or some blocks are damaged, there are several ways to recover the lost area by using the existing pixels. The spatial correlation techniques, in general, conceal the lost area by estimating it from the previously received or successfully concealed neighboring pixels. Usually the decision of which blocks, or pixels around the lost blocks are to be chosen is made by using some boundary check to minimize the blocking effects and obtain a natural view. Since to conceal a lost area with spatial concealment techniques, the frame needs to be at least partly decoded, whole frame losses are not possible to recover. Whereas the temporal concealment techniques use the temporal correlations of already encoded previous or next frames to estimate the lost area. Therefore whole frame losses can be concealed with temporal techniques. Unfortunately the temporal concealment techniques do not perform well when the scene contains fast movements. The hybrid concealment techniques combine spatial and temporal techniques.

In [8], the error concealment techniques for previous 15 years of 1998 are reviewed. In the paper, the proposed algorithms are grouped in five classes where either the DCT coefficients or the pixel values or the MV values and the coding modes are predicted. Those are motion-compensated temporal prediction, maximally smooth recovery, projections on to convex sets (POCS), spatial and frequency domain interpolation and recovery of motion vectors and coding modes techniques.

In motion-compensated temporal prediction method, the lost MB is replaced with spatially corresponding MB in the previous frame or compensated with the motion vector if not lost. In this method a major assumption is made, where the video is coded and transmitted in a layered structure. The low frequency components of DCT coefficients and motion vectors are coded together, named as base layer, whereas the high frequency components of the DCT coefficients are coded and transmitted separately, named as enhancement layer. When the enhancement layer is lost, Kieu et. al. proposed to use the MVs of the base layer to estimate the high frequency components from the previous frame [9]. When the base layer is received correctly, using the MVs, the corresponding MB in the previous frame is found. Then by taking the DCT of the MB, the DCT coefficients are obtained. Using the high frequency components of the compensated MB from the previous frame is claimed to improve the visual quality instead of setting the lost coefficients to zero.

In maximally smooth recovery method, smoothness property of the most image and video signals is used. The missing DCT coefficients are estimated to minimize the temporal and spatial variation between the concealed block and the temporally and spatially neighboring blocks. The smoothness constraint is given in [10] as minimization of the boundary difference between the neighboring blocks. When the DC coefficients are lost, then minimizing only the difference between the perimeter of the lost MB and surrounding pixels gives good results. However when the higher coefficients are lost, then minimizing the difference between each pixel and the neighboring pixel being closest to the boundary of the MB is needed. The basic smoothing schemes are given in the Figure 3-1. With the smoothing criterion, there are several possible estimation techniques for the lost block. With the spatial or temporal techniques, the pixels, MVs or the DC coefficients are estimated iteratively and the smoothness constraint is tried to be minimized.



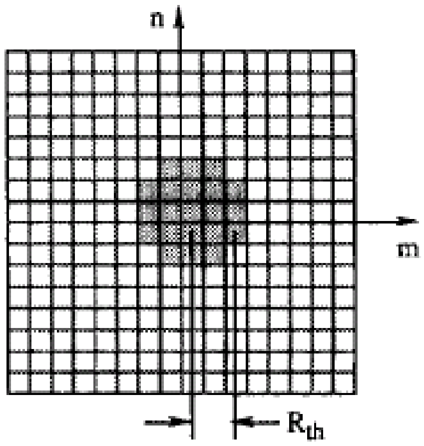
(a) (b)

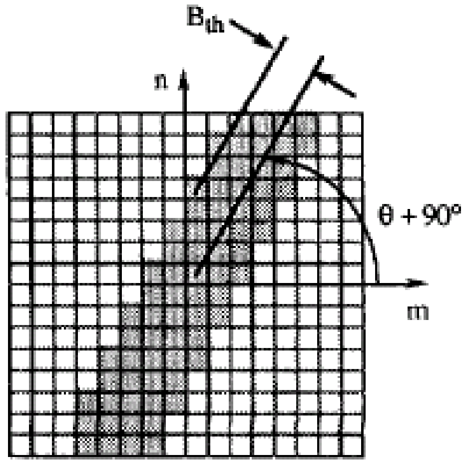
Figure 3-1: The smoothing constraints [10] where (a) only the difference between the surrounding pixels and adjacent pixels is minimized, and (b) the difference between each sample and its neighboring sample which is closer to the perimeter is minimized.

Since the smoothing constraint given above considers omni-directional boundary matching, it causes the edges to be blurred. As given in [11] and [12], the smoothing constraint is improved to handle the edges in the lost MB. The smoothing constraint is modified such as directional boundary matching criteria are added to the algorithm.

In projections onto convex sets (POCS) method, as given in [13], firstly the lost MB is categorized as an edge or a smooth block by using the neighboring pixels. Then if the block is an edge block, the direction is estimated from the neighboring blocks too. The lost MB and the neighboring blocks are combined to have a large block. Then this large block transformed and filtered. The filter is a low pass if the lost

MB is predicted to be smooth and an edge directional band pass filter if the lost MB is predicted to has an edge. An illustration for the filters is given below.





(a) (b)

Figure 3-2: The filters for POCS method (a) low pass filter, (b) edge directed band pass filter

By filtering the coefficients, if the MB is smooth, the high frequency components are eliminated. If the MB has an edge, the frequency components related to the edge are reserved and remaining coefficients are filtered. After filtering, the coefficients are truncated. Then the resulting coefficients inverse transformed and the lost MB is taken for the next iteration. The process is applied until the lost area remains unchanged after those operations. A block scheme is given in Figure 3-3.

Later Robie and Mersereau proposed to use Hough transform to determine the directionality of the lost MB [14]. First the image is segmented as dark foreground, light foreground, and background. Then using Hough transform, best direction for interpolation or filtering is determined.

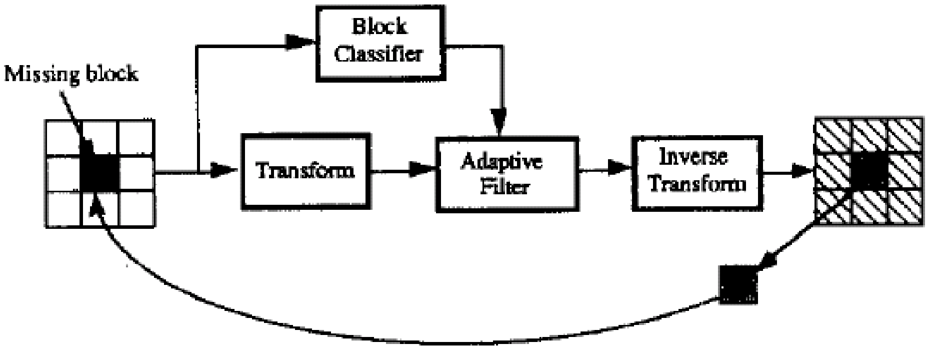
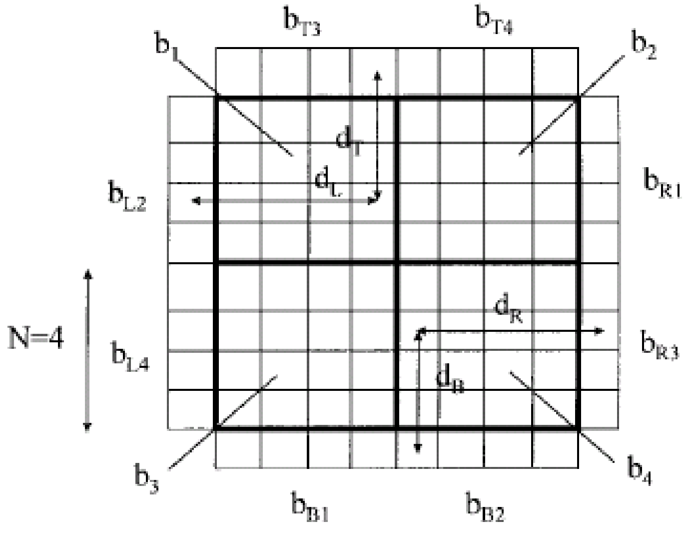


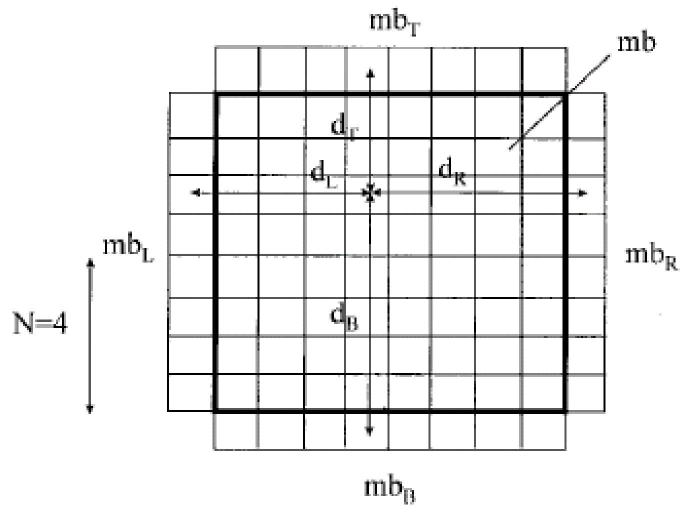
Figure 3-3: Iterative POCS method

In the spatial- and frequency-domain interpolation, the DC coefficients or the pixels are interpolated according to the four neighboring MBs. In [15], a simple interpolation to find DC coefficients is proposed. The DC coefficient of a lost MB is found by a simple weighted average of the DC coefficients of 4 neighboring MBs. More generally, in [16], Hemmami et. al. proposed an iterative linear interpolation technique which can be applied in frequency domain and spatial domain both. The DCT coefficients or the pixels of a lost block are concealed by taking a weighted sum of the four neighboring coefficient or pixel arrays. In the summation the lost block itself is summed too. In the beginning it is assumed to be zero, and former results are input for the next iterations. The weights should be selected to maximize the smoothness in the boundary of the lost block.

In [17], the interpolation is done using only the neighboring pixels instead of the whole neighboring blocks. A lost MB is concealed pixel by pixel with an average of 4 neighboring pixels located in the same horizontal or vertical positions. However, if a block is to be concealed, the two neighboring pixels are used in the same way. The MB and block interpolations are shown in Figure 3-4.



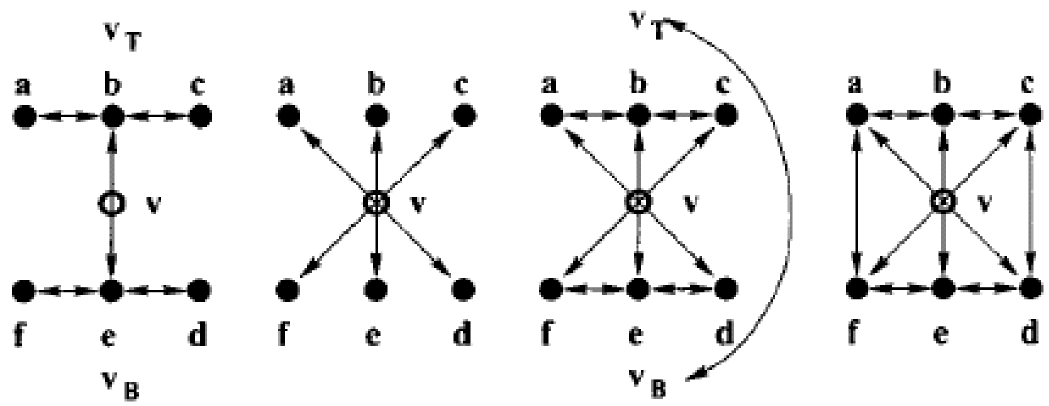
(a)



(b)

Figure 3-4: (a) Block based, and (b) MB based interpolation [17].

In [17], an error concealment algorithm is given when the mode and the motion vectors are lost. Based on the assumption on spatial and temporal smoothness property of the video sequences, the mode and the motion vectors are interpolated from the neighboring blocks. Four different methods for choosing MVs given in the paper are; setting zero by default, using previous corresponding MV, using average of spatially neighboring MVs, using the median of the spatially neighboring MVs. When a packet is lost, usually the left and right neighboring MBs are already lost. To overcome the problem, in [18], a spatial interpolation method is proposed to utilize only top and bottom neighboring MBs.



(a) (b) (c) (d)

• : Adjacent MV   o : Lost MV

Figure 3-5: The interpolation scheme employed in vector rational interpolation technique; (a) 2-stage 1-D, (b) 2-D, (c) 2-stage combined 1-D and 2-D (d) 2-D in all directions.

In the Figure 3-5, the four interpolation schemes proposed in [18] are given. The interpolation in each case is done in 1 stage or 2 stages by summing and weighting the elements connected with the arrows. Then the weighted sums are weighted and averaged to obtain either V directly, or in 2 stage cases, VT and VB. In the two stage cases, V is obtained by taking the average of VT and VB.

The temporal concealment methods propose to replace a missing block with another block placed in previous or next frames. A major problem exists for those methods because assuming simple block movements does not perform well in case of splitting, rotation, zooming and so on. Simple block translations cause artifacts at the borders. In [19], Atzori et. al. proposed spatio-temporal concealment technique using boundary matching algorithm and mesh-based warping technique to conceal a block in case of non-uniform translational movement. To conceal the block, first using a boundary matching algorithm a candidate for the lost block is found from the previous frame. This boundary check algorithm applies smoothness properties as mentioned before. Then using mesh based warping, the block is partitioned to meshes, and nodes are placed. And by using the neighboring pixels, movements of the nodes are predicted. The meshes are affine transformed accordingly. The block scheme for the proposed algorithm is given below.

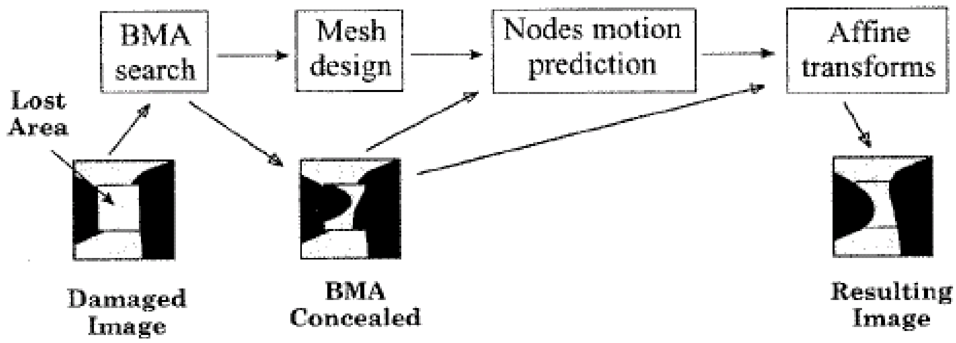
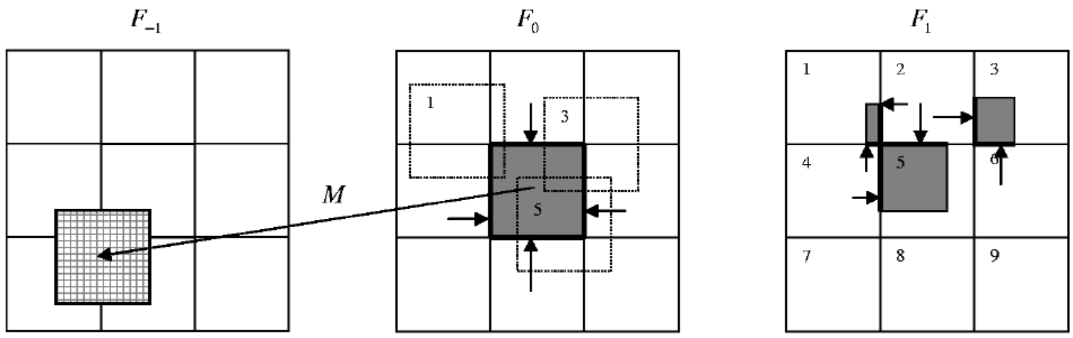


Figure 3-6: BMA and mesh based warping algorithm [19]

Motion compensation used in video coding standards reduces the temporal redundancy in the coded bit stream. However it also causes the error to propagate through the frames that reference an erroneous frame. Error concealment schemes consider a smoothness constraint between the borders of a concealed block and the surrounding blocks or pixels. Unless the concealment technique perfectly recovers the lost block, some error will propagate to the frames those references the concealed block too. Lee et. al. Proposed multiframe error concealment technique to reduce the effect of erroneously concealed block in the succeeding frames [20]. In this technique, in the first phase with either a spatial or a temporal method, a candidate block is chosen from the reference frame. The boundary smoothness error is calculated. Then, in the second phase, the blocks that references any part of the candidate block are determined. After motion compensation, the boundaries which are copied from the candidate block are subject to a smoothness constraint. The two smoothness errors are summed. Iteratively repeating the two phases, by minimizing the sum of two smoothness error, concealed block is determined. The following figure summarizes the error concealment process. In F_0, the grey area is the lost MB. The dashed blocks in F_0 are the references for the same numbered blocks in F_1. The block in F_{-1} that maximizes the borders shown with arrows in both F_0 and F_1 is chosen.



(a) (b) (c)

Figure 3-7: Multiframe error concealment technique [20].

In low bit rate systems, mostly a packet carries the information to decode a large part of the frame. Therefore when a packet loss occurs, the blocks that surrounds the blocks to be concealed is also lost, and block concealment techniques fail. Chen et. al. propose a temporal concealment algorithm to recover the lost block recursively [21]. In the paper first a block concealment scheme is introduced where lower and upper neighbors are not lost. Then using the same algorithm, more than one line of MBs can be concealed. In the Figure 3-8, E is the lost block, E_l is the lover half of the lost block, and E_u is the upper side of the lost block. B_{TOP} and B_{BOTTOM} are the

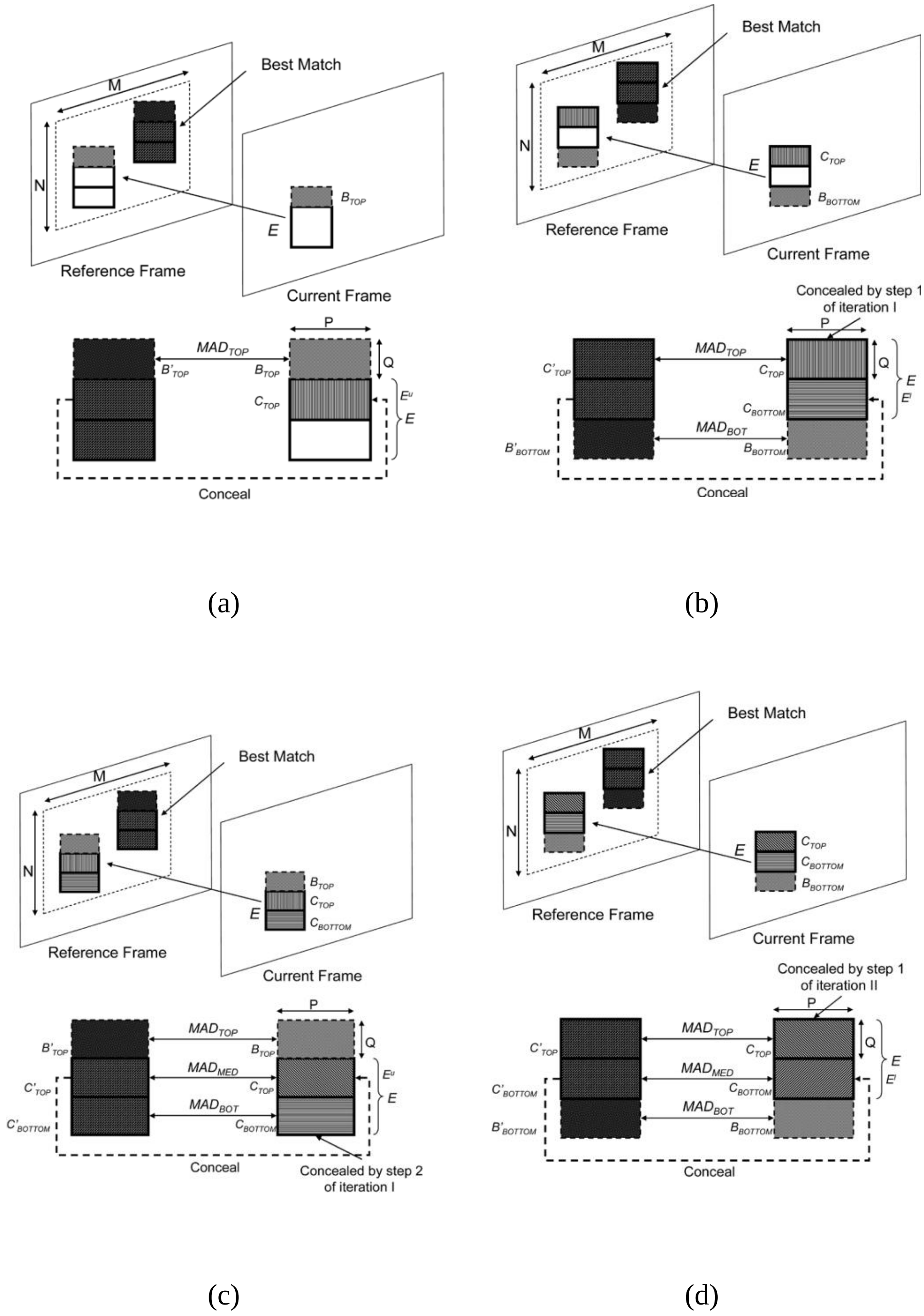
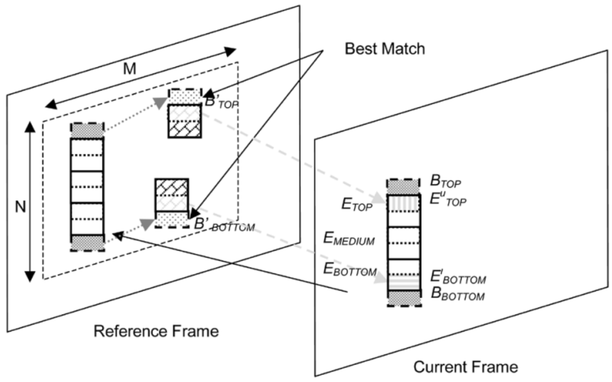


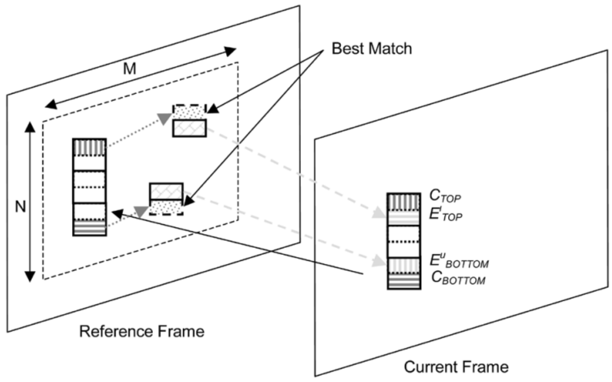
Figure 3-8: Recursive error concealment (a) first step of first iteration, (b) second step of first iteration, (c) first step of second iteration, (d) second step of second iteration.

neighboring half MBs of the lost block E. B'_{TOP} and B'_{BOTTOM} are the matching half blocks for B_{TOP} and B_{BOTTOM}. C_{TOP} and C_{BOTTOM} are the selected half blocks for the E_u and E_l respectively. C'_{TOP} and C'_{BOTTOM} are the candidate half blocks from the reference frame. In the first step of the first iteration of the algorithm as shown in (a) of Figure 3-8, a candidate upper half block, C'_{TOP}, for the E_u is found minimizing the mean absolute difference (MAD) between B_{TOP} and B'_{TOP}. The selected half block C_{TOP} is used in the second step of the iteration. A candidate lower half block, C'_{BOTTOM}, for E_l is found using C_{TOP}-C'_{TOP} and B_{BOTTOM}-B'_{BOTTOM} as in (b) of Figure 3-8. C_{BOTTOM} is saved for later iterations. In the later iterations a weighted sum of 3 MADs is minimized to update a lower or upper part of the lost block. In the first step of the second iteration, C_{TOP} is found using B_{TOP}-B'_{TOP}, C_{BOTTOM}-C'_{BOTTOM} and C_{TOP}-C'_{TOP} as shown in (c) of Figure 3-8. Then in the second step of the iteration, minimizing B_{BOTTOM}-B'_{BOTTOM}, C_{BOTTOM}-C'_{BOTTOM} and C_{TOP}-C'_{TOP} differences, C_{BOTTOM} is found as shown in (d) of Figure 3-8.

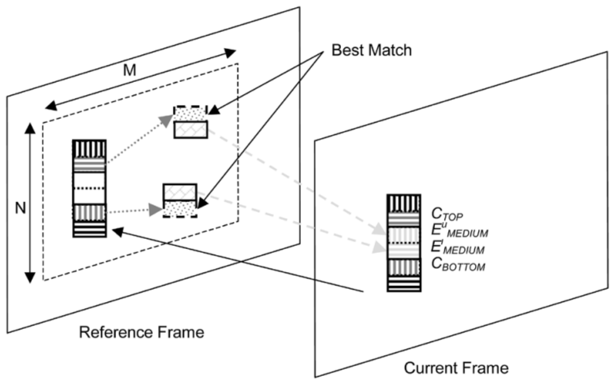
The iteration 2 of the algorithm is repeated until the sum of weighted MADs for both step 1 and step 2 are not changed, meaning the lost block converges, or a threshold for iteration number is exceeded. For the case where top or bottom MBs are not available, the first step cannot be passed. When the top or the bottom blocks are not at hand, the number of the steps in iterations is increased. An illustration is given in Figure 3-9. To conceal the lost block, in the first step, upper most and lower most half blocks are concealed. Then having the upper most and lower most half blocks concealed, the method is applied again to obtain the lost area towards the middle lost block. When the middle block is concealed, as explained before, step 1 and step 2 of iteration 2 is applied and the concealment is tuned.



(a)



(b)



(c)

Figure 3-9: Recursive error concealment when top and bottom MBs are not available. (a) step A, (b) step B, (c) step C

Recursive error concealment method gives a good solution for group of MBs losses, however, still at least top most and bottom most MBs are needed to be available. In [22], a concealment scheme is proposed for whole frame losses, where no neighboring MBs or MVs exist which can be used in concealment. Belfiore et al. proposed to estimate a lost frame from the optical flow. With this method firstly an optical flow is estimated for the pixels of the last frame, using MV information of few past frames. Then the last frame is projected using the optical flow information. The areas which are covered more than once are averaged, and the areas which are not covered with any projection are recovered by interpolation. The method performs well however estimating optical flow for all pixels and median filtering gives extra computational load.

Another method is using past MVs estimating current MVs instead of getting optical flow estimation to recover a whole frame loss. In MV extrapolation method, the MVs of the lost frame are extrapolated from the last decoded frame [23]. Then the extrapolated MVs are compensated with the last frame and lost frame is concealed. In this method MVs are extrapolated for 8x8 blocks. For any MV overlapping to the 8x8 block, the block is assumed to have extrapolated MV and compensated with this value. This causes block artifacts. In [24], pixel based MV extrapolation is proposed to overcome the artifacts. In this method each pixel is concealed independently with MVs extrapolated from the previous frame. The MVs of the past frame are projected to the current frame. When a pixel is covered by more than one MV, it is compensated with average of the MVs. When no MV covers the pixel, then this pixel is compensated with the MV of the block which spatially corresponds to the uncovered pixel. This method cannot give better performances in large motion cases because of the direct MV copying in the uncovered areas. Later, Zhou et al. used MV extrapolation with Lagrange Interpolation to improve the algorithm [25].

3.3 ERROR CONCEALMENT FOR STEREO VIDEO

In monoscopic video sequences only redundancy exists in the spatial and temporal domain. However in the stereoscopic video sequences, there is interview correlation. Interview redundancy is used in coding standards to compress the stereo video more efficiently. It can be used for the recovery process in case of any error along with the spatial and temporal redundancy.

In [26], a concealment technique is proposed for the schemes where the views of a stereo pair are coded separately and interview redundancy is not utilized. In this algorithm, first step is to extract some feature points around the lost block using gradient measurement. In the second phase, matching feature points in the other view are found using epipolar geometry. In the Figure 3-10, the matching feature point process is shown.

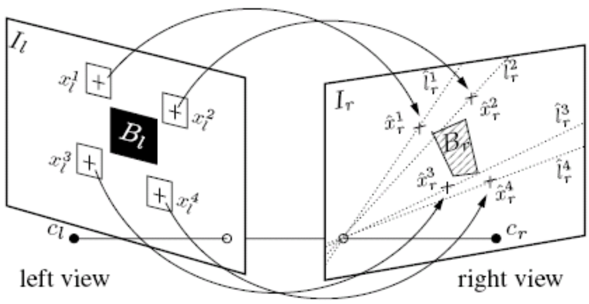


Figure 3-10: Matching the feature points in the two view of stereo video

Bl is the lost area; x_l and x_r are the feature points. l_rs are the epipolar lines passing through the feature points. In the next stage with Random Sampling Consensus (RANSAC) algorithm, the transform coefficients are found coarsely. Then using Newton method, finer transform coefficients are found which minimizes the sum of errors between boundary pixels neighboring the lost block and the corresponding pixels in the other view. In the next stage, if the disparities of the matching points are varying above a threshold, then a 3D block smoothing is utilized. This 3D block smoothing is a weighted average of the boundary pixels and the reference pixel from the other view. The flow chart for the proposed algorithm is given in Figure 3-11. It is stated in the paper that smoothing increases PSNR quality however decreases the visual quality. The algorithm outperforms the classical mono concealment techniques however requires high computational power for the estimation of transform coefficients between the two views. Also the method does not provide solution for whole frame losses.

Another coding scheme for the stereo video is to code the views by using interview correlations. In those schemes the MBs in a view can either be motion compensated or disparity compensated. In [27] an algorithm for full right frame losses is proposed. In this algorithm firstly a relativity analysis is made to decide either temporal or interview correlation technique is to be applied. Relativity analysis is the comparison of the numbers of motion compensated MBs and disparity compensated MBs. In a sequence, if the number of MV compensated MBs are higher, than the lost right frames are concealed with the motion extrapolation technique as in the mono cases. If the number of DV compensated MBs are higher, then the DVs used to decode the last successfully received right frame are used. The algorithm performs well. However it is not preferred to conceal the frames with only one technique with the information of the sequence being either MV or DV predicted. This causes the concealment technique not to be precise in long sequences.

In [28], disparity between last right and left frames are estimated. The pixel-wise disparity is found by minimizing the sum of absolute difference for a pixel and a pixel window surrounding it. Then the disparity vectors are median filtered. In the second step the pixels are categorized as whether they are in an occlusion area or not. To identify, the uniqueness property is used such that, if a pixel corresponds to more than one pixel in the other view, than there exists an occlusion. Non-occluded pixels are concealed by projecting the MVs of the left sequence with disparity vectors. Occluded pixels are directly copied from the last right frame if neighboring areas has small temporal replacement. They are copied from the closest neighboring pixel otherwise.

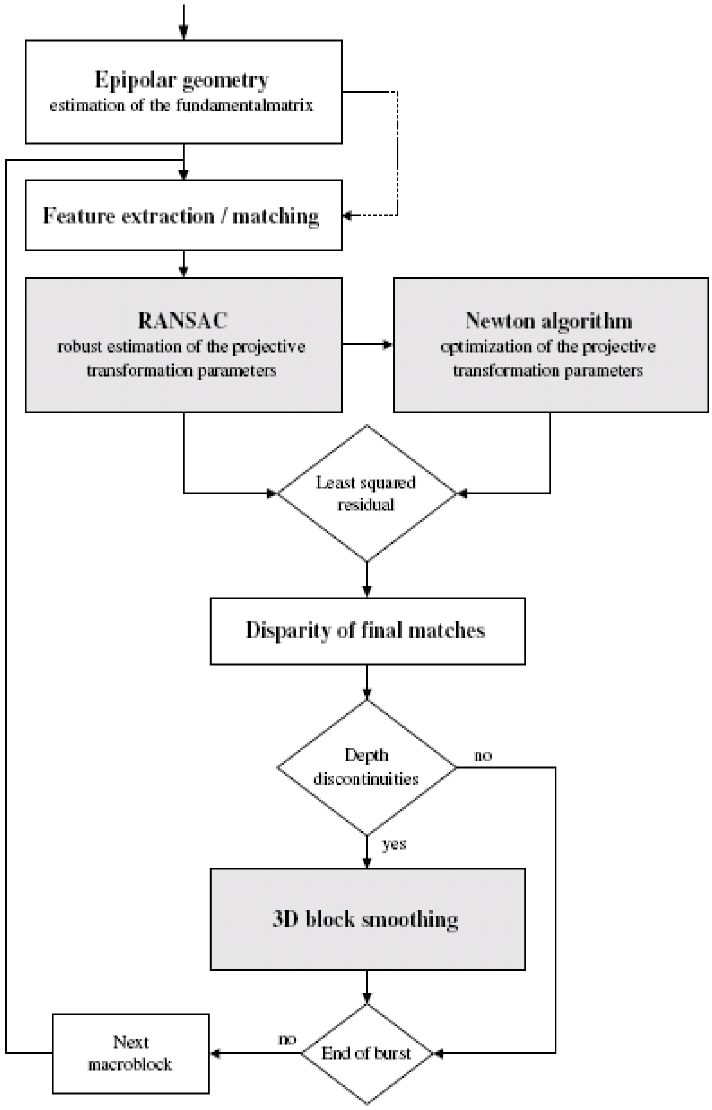


Figure 3-11: Flow chart for of proposed algorithm in [26].

In [29], a subjective evaluation test is made to compare the effectiveness of three simple concealment techniques. First one is frame freeze where the last frame is copied in place of the lost frame. The second one is double freeze where in case of any loss; the two views are copied from the last two frames of the two views. The third one is base copy, where in case of any loss; the other view is copied in the place of the lost frame. In the third method, 3D vision is replaced with the 2D vision. From the subjective test conducted with 38 participants, it is deduced that copying last frame for only lost image is preferred for low disparity sequences and for low probability of error. However it is preferred to switch from 3D to 2D in case of error bursts and high disparity.

In [30], Yan proposed a motion vector recovery algorithm for 3D transmission. The algorithm is proposed for depth plus video systems. Depth and video are correlated such that MBs, having similar depth values in the depth frame, have similar MV values in the video frame. For the similar depth values in the depth image most likely to belong to the same object, therefore those MBs have the same motion. Using this property, the algorithm utilizes block matching and selects a MV among the ones having similar depth values in a search range. This makes the algorithm computationally less complex compared to the ordinary BMA algorithms. In [31], Liu et al. introduced a similar approach to conceal the video frames. The lost MB is specified as a smooth or edge MB by histogramming the depth image. Then the block is concealed with the traditional temporal error concealment methods if it is smooth. If the block contains edges, then it is partitioned as foreground and background areas. Then each part is concealed with temporal error concealment methods, having MVs of the neighboring areas. In [32], Ali et al. proposed a similar algorithm however the foreground and background areas are concealed with spatial techniques.

In [33] the correlation between depth and video is used in a different manner. Since depth and video have similar motion characteristics, when a depth frame is lost, it is compensated with the video MVs. In this paper, using color video sequence's MVs for concealment is investigated only. The algorithm is not given for the color video frame losses.

In [34], Hewage et al. proposed to recover color video frames using depth MVs as recovering depth frames from color video MVs. In this algorithm, unlike the other concealment techniques, modifications are to be made in the encoder side. When encoding the depth video, the motion estimation part is disabled. Instead of motion estimation, directly the color video MVs are used, and residuals are transform coded. The two bit stream has the same MVs with this method. Therefore in case of any loss in any view, the MVs of the other view can be used. This method performs very well. However, the bit stream size for depth view is increased. Also modifying encoder makes the method more like a forward error correction algorithm. The method cannot be implemented in the receiving systems only.

In this thesis, the algorithms requiring large memory sizes or high computational power are avoided i.e motion vector interpolation, block matching, projection and so on. The error concealment methods that are implemented in this thesis are intended to be used on mobile platforms. Also the iterative methods which may not converge in a reasonable time interval are not suitable for realtime systems having limited computational power. In the next chapter, low cost error concealment algorithms that are implemented in this study are given.

CHAPTER 4

IMPLEMENTED ERROR CONCEALMENT METHODS

4.1 ERROR CONCEALMENT FOR 2D+DEPTH CODING

In 2D+Depth coding, as stated before a depth map, which is a monochromatic video scene, and the colored video is encoded and transmitted. In the receiver the colored video is directly used as the left view. The depth map and video frame are used to synthesize the right frame. In this part of the thesis, two error concealment techniques for 2D+depth are implemented. First a simple frame copy method and then a complementary motion vector usage method for 2D+depth are implemented.

4.1.1 FRAME COPY BEFORE RENDERING

In 2D+depth, when a depth frame or video frame is lost, right frame cannot be rendered. Frame copy is a simple concealment technique when a frame is lost. However, directly copying the last right frame to the place of current lost right frame discards some information unless the two views are lost at the same time. If at least one of the views is received correctly, this information can be used for the current scene synthesis. In the frame copy before rendering (FCBR) method, instead of copying last rendered right frame, current correctly received frames are used. The block scheme of the algorithm is given in Figure 4-1.

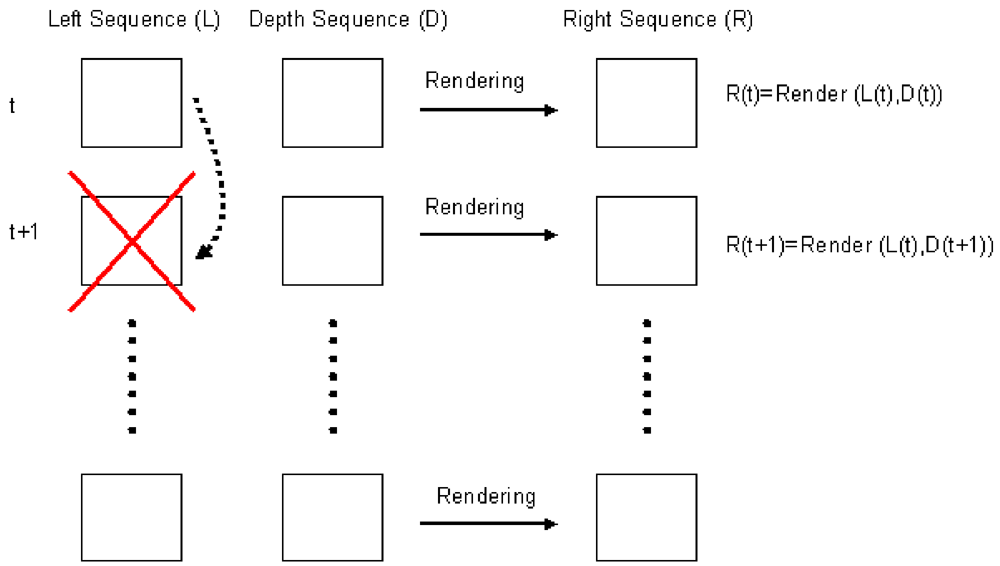


Figure 4-1: Frame copy before rendering method when a left frame is lost

In the figure, sample concealment implementation for the left frame loss is shown. If a left frame is lost at time instant t+1, by using the depth frame at time instant t+1 and left frame at time instant t, the right frame at time instant t+1 is rendered. If a depth frame is lost similar approach is used. The method is independent of the coding-decoding structure because the concealment is done in the pixel domain, where decoder already gives out the successfully received frames. For the performance analysis, for each frame the concealment technique is applied. The simulations are made for both depth frame losses and left frame losses.

4.1.2 COMPLEMENTARY MOTION VECTOR COPY

The two sequences, depth and video, are highly correlated in a 2D+Depth system. As can be seen from the figure, the motions of the objects in the videos are parallel to each other. This redundancy of the motion in the two sequences can be used to for concealment in case of a frame loss.

Figure 4-2: 268th (up) and 269th (down) frames of the depth(right) and video (left) sequences of Rollerblade video.

In the complementary MV usage method, when a frame is lost the correlation between the depth and color video sequences is used. Since the motion is parallel in the two sequences, in case of a loss, the motion can be estimated from the other sequence. In the method the depth video is complementary for the color video and the color video is complementary for depth video from the aspect of similar MVs. For the implementation, jm17.2, which is an H.264/AVC encoder/decoder, is used. The concealment technique is implemented for the simulcast coding structure where the two video is coded such as two separate monoscopic videos are encoded. The structure is given in Figure 4-3.

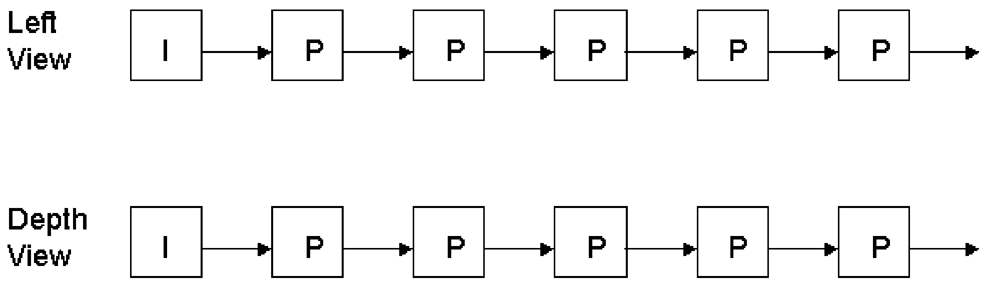


Figure 4-3: Simulcast 2D+Depth coding structure.

In the motion compensation phase of jm17.2 decoder, a buffer is constructed for the MVs. For each successfully decoded block, the MVs are saved to be used later. According to H.264, maximum of 16 MVs can exist in an inter coded 16x16 MB, where the MB consists of 4x4 blocks. Therefore, for each MB of the frame, 4 by 4 MV array is needed. For whole frame, width/4*height/4 sized MV buffer is constructed. If a MB is encoded with 4x4 blocks only, each MB is mapped to one buffer index. However if the blocks are larger, i.e. 8x8, 16x8, 8x16 or 16x16, then the block is mapped to multiple indices to cover the corresponding place in the buffer. Since only inter coded blocks of P frames has motion vector data, the intra coded blocks does not contribute to the motion vector buffer. Those indices of the buffer are filled with zeros. Therefore, for the empty indices, zero motion is assumed. The figures show how the buffer is filled with successfully received motion vectors. Figure 4-4 shows MB level mapping and Figure 4-5 shows an example 16x8 block mapping in the MV buffer.

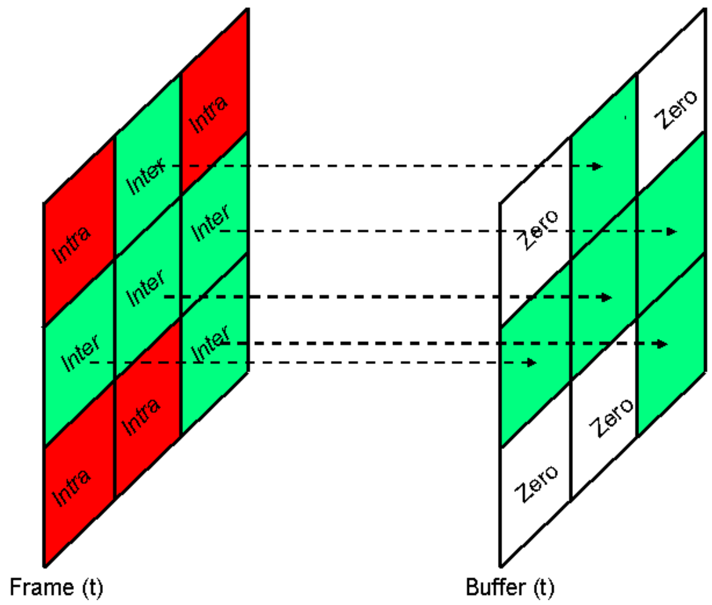


Figure 4-4: Construction of MV Buffer (MB Mapping)

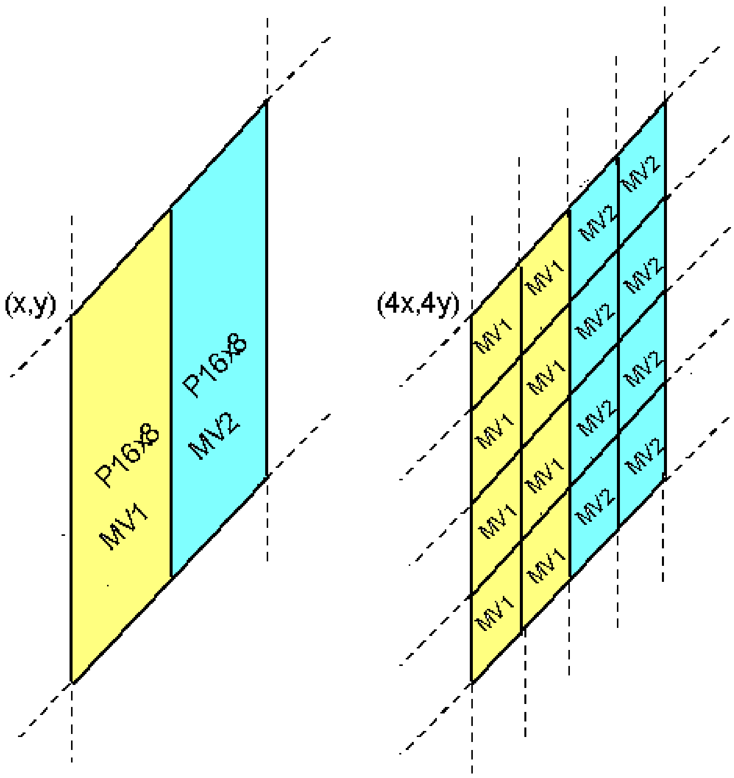


Figure 4-5: Construction of MV Buffer (Block Mapping)

When a packet is not received correctly, a slice is lost in a H.264/AVC decoder and the buffer is used for concealment. Although the method can be used to conceal part of the frame, in this thesis whole frame losses are considered. Using the complementary sequence’s buffer, the frame is decoded as if the MVs are correctly received. For the lost frame, firstly imaginary 4 by 4 blocks are generated as if all the frame is encoded with 4 by 4 blocks. Then each block is decoded with MVs taken from the complementary MV buffer. Here the residuals are assumed to be zero while decoding. A block schema for the implementation is shown below. As seen from the Figure 4-6, to conceal the lost frame, last successfully decoded frame is used as reference frame for motion compensation. Complex referencing and flexible ordering schemes are not considered for the implementation.

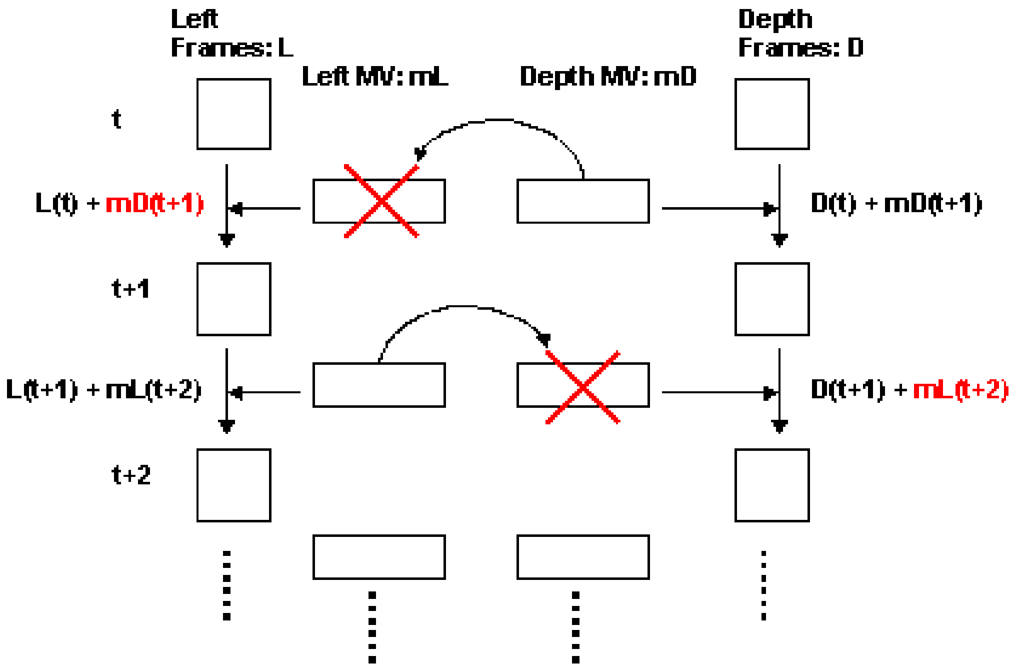
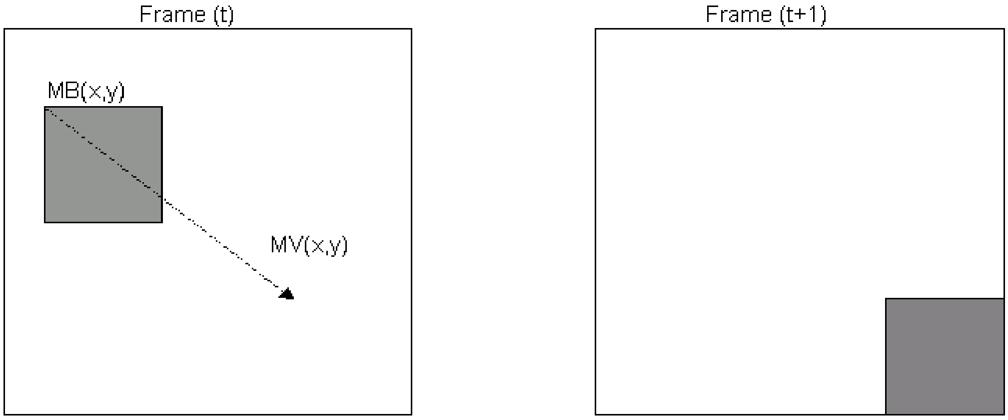
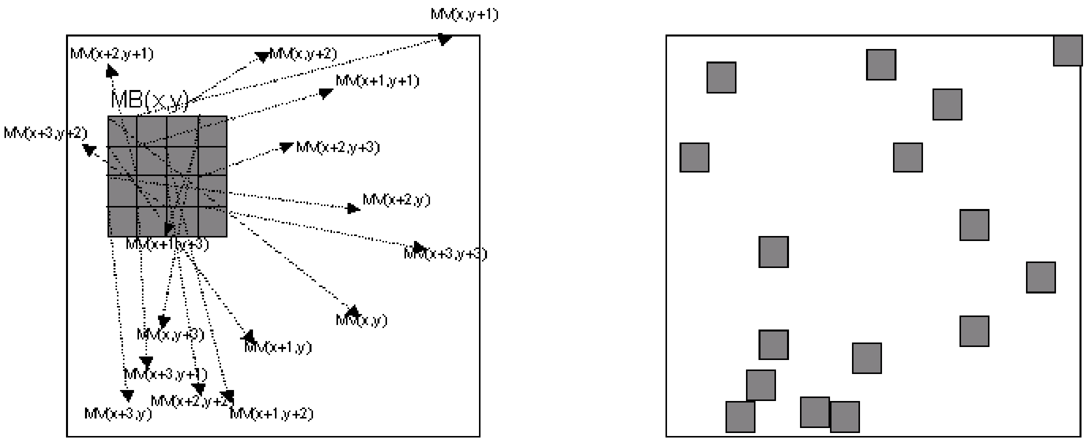


Figure 4-6: Complementary Motion Vector Usage Method for 2D+Depth.

Although 16x16 block decoding might decrease the computational cost and memory usage, 4x4 blocks are generated to increase the motion resolution. A 16x16 block decoding process use only 1 of each 16 MV in a MB. Via 4x4 block compensation, the 16x16 MB can be splitted and the error can be reduced. This utilizes the use of each individual index of the buffer. An extreme example of how severe the effect of using 16x16 block decoding can be is shown in the Figure 4-7. If the MB is decoded as a 16x16 block, then the whole MB is transferred. However when the MB is decoded as 4x4 blocks having individual MVs, each block can be transferred to an individual place.



(a)



(b)

Figure 4-7: (a) 16x16 block motion compensation, (b) 4x4 block motion compensation

4.2 ERROR CONCEALMENT FOR STEREO VIDEO

The two views of a stereoscopic video are highly correlated. Therefore in case of a loss, the redundancy between the two views of stereoscopic video can be used. In this part of the thesis, two error concealment techniques for stereo video are implemented. First complementary motion vector usage method for stereo video and then a last disparity vector usage method are implemented.

4.2.1 COMPLEMENTARY MOTION VECTOR COPY

In the stereoscopic video, although there is a shift between the two views, there is a parallelism between the motions of the two views. The shift between the views causes the motion vectors at the edges of the objects and the occluded areas to be different. An illustration is given in Figure 4-8. However for the large objects, and background movements, the two views almost have the same motion characteristics.

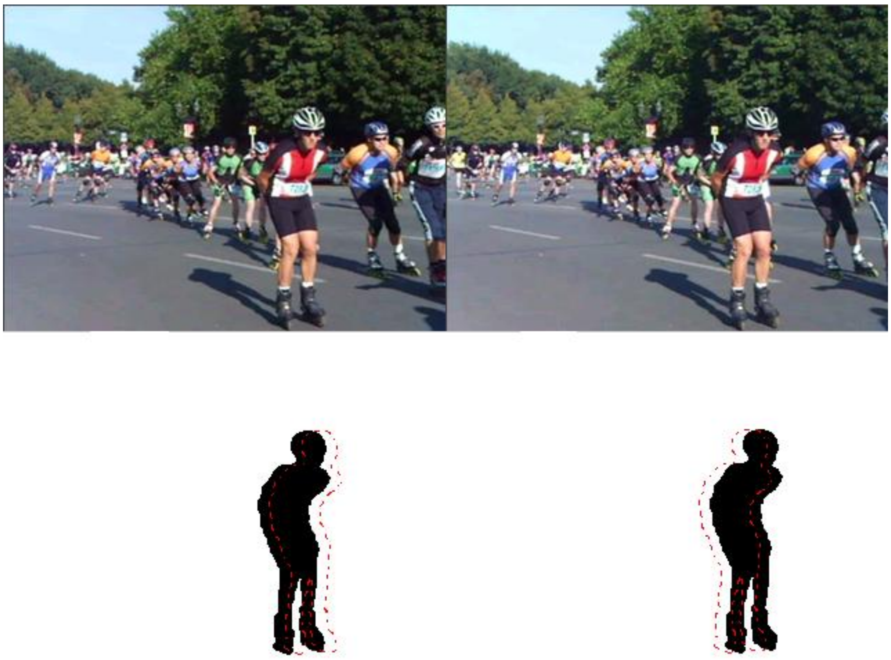


Figure 4-8: Shift and occlusion in the stereoscopic video (dashed lines are the corresponding object edges in the complementary view)

For the simple systems such as mobile devices which use low resolution video, the error caused by the shift may be ignored. In this method, the same procedure as in CMVC of 2D+Depth is applied to the stereoscopic videos. With this method assuming the motion vector of the object is unique, the intersecting areas of the moving objects in the two views are concealed with very good quality. The background areas will also mostly be concealed with very little distortions, since the background movements are either small motions or due to camera movements where the whole background has similar MVs. Therefore the shift does not affect the unique and small background MVs too.

The method is implemented for simulcast stereo coding structure where the two sequence coded-transmitted-decoded independently. The simulcast coding structure for stereoscopic video is shown below.

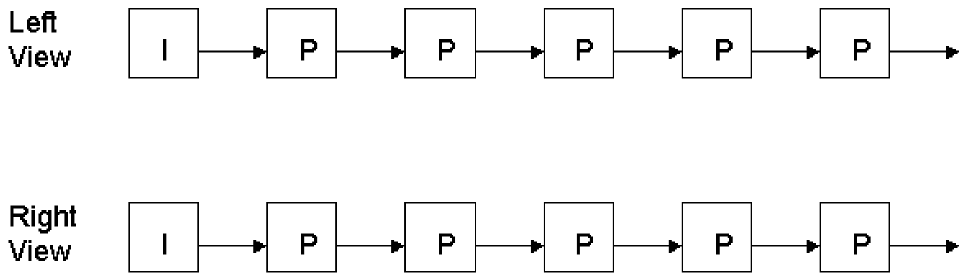


Figure 4-9: Simulcast Coding Structure for Stereo Video

As in the CMVC of 2D+depth method, when a packet is considered to be lost, then instead of the lost frame, imaginary MBs are constructed and decoded. The lost residuals are not tried to be recovered, and the motion vectors are directly taken from the complementary view (if the lost frame is left, then complementary is right sequence, and vice versa). The MV mapping is exactly the same as explained in 4.1.2, so it is not explained here again. The block schema of the complementary motion vector usage is given in Figure 4-10.

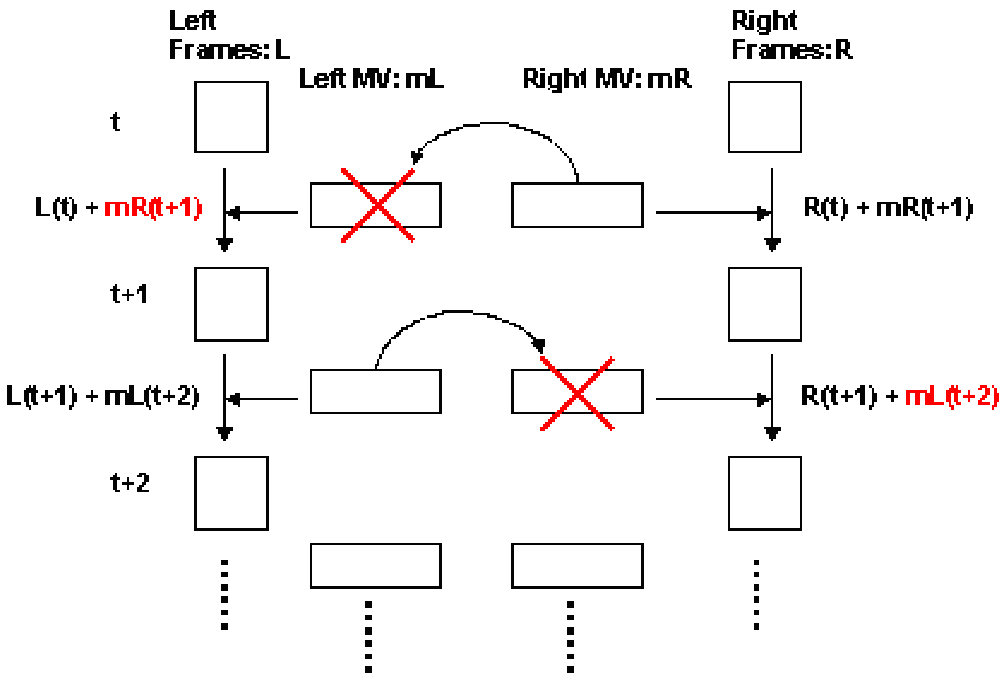


Figure 4-10: Complementary Motion Vector Usage For Stereoscopic Video

Since the implemented methods are intended to be run on mobile platforms, discarding the errors at the edges and occluded areas not to increase the complexity of the calculations is preferred. The MBs of the lost frame is decoded as 4x4 blocks as in 2D+depth case.

4.2.2 DISPARITY VECTOR COPY

The two views of a stereoscopic video are very similar to each other, except for a shift (displacement between views) and occluded areas. Although the motion vectors change fast in a fast scene, the shift remains nearly the same throughout the video because the interview angle and the distance of the cameras remained constant. This property of stereoscopic video enables to conceal a right frame by shifting the left frame exactly the same amount as the last displacement between the two views. The method is proposed for the simplified MVC coding structure given below.

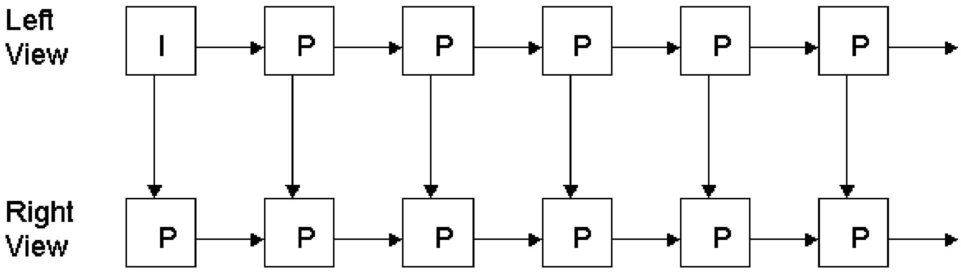


Figure 4-11: MVC coding structure

However the implementation is first done as given in the Figure 4-12 to investigate the effect of using last Disparity Vector more appropriately. In this custom structure, the left frames are P coded as usual. However the right view is coded with DV compensated inter blocks, and intra blocks. This means no motion vector is used for right view, so the effect of using last DVs is magnified.

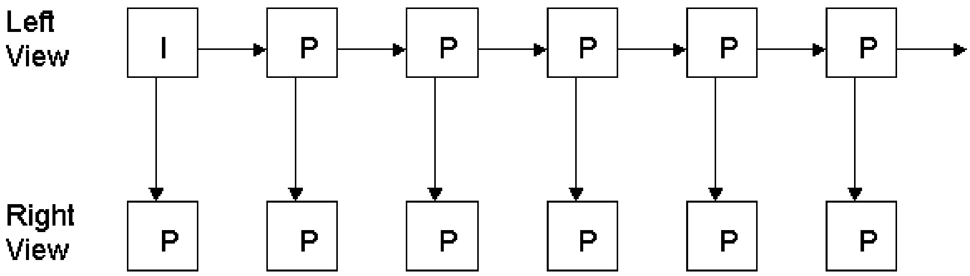


Figure 4-12: Modified MVC coding structure for last DV usage method implementation

With this structure, when a right frame is lost, last used DVs are used to conceal current right frame. In the compensation phase of the jm17.2 each correctly received DVs are saved to a DV buffer. The DV mapping is constructed just the same as in the complementary motion vector usage in 2D+depth. For each 4x4 block, a DV is saved. When a right frame is lost, from the DV buffer the last DVs are taken. The imaginary blocks are constructed and with the last DV values the blocks are compensated. The current correctly received left frames are used as reference in the motion compensation phase. The schema of the method is given below.

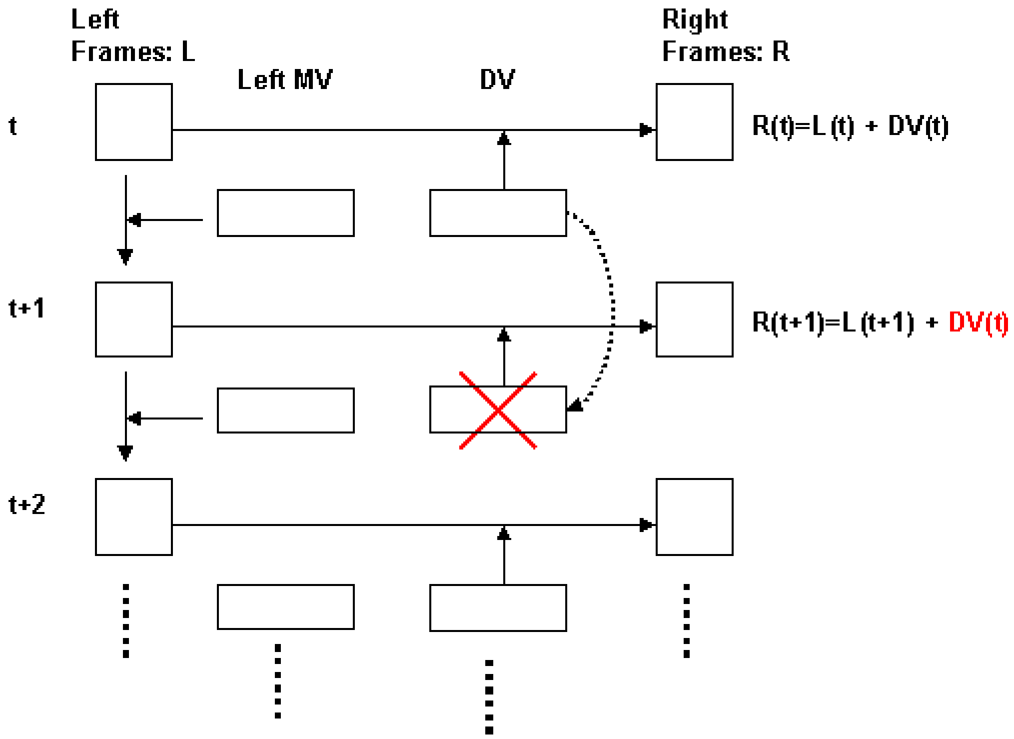


Figure 4-13: Last DV usage method (for modified MVC structure)

If the last right frame was coded with inter P blocks only, there would be no need for the intra filling. The entire buffer would be filled with the last DVs. However, in a usual coding scheme, the frames are coded with intra and inter blocks. Therefore the buffer cannot be filled with only DVs. In the motion vector copy case, filling the empty indices of the buffer with zeros corresponds to the simple block copy, or zero motion. However in the last DV copy method, compensating the left frame with zero DVs means a motion which is mostly negative of the interview shift/displacement. To simulate the zero motion copy in the former method, for the intra coded blocks in the former right frame, a default shift, instead of zero DV, should be used. Assuming there is no motion in the video, and all the objects have the same distance from the two camera, and all the blocks are coded with DVs, it can be said that, there would be a unique DV. In this case the right frame is just a shifted version of the left frame. Therefore to estimate the DVs for the intra coded blocks, a default shift is assumed between the views. This default shift is calculated with histogramming the x component of non-intra coded blocks. The most frequent DVs are (if more than one, they are averaged) used for the intra blocks. The histogramming is done in the x direction only, because the stereo camera setting is assumed to be horizontal by default. The process is shown in the figure below.

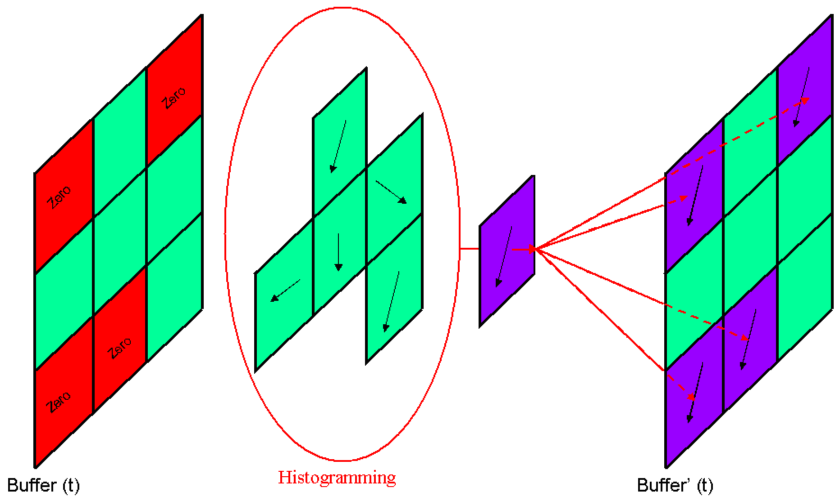


Figure 4-14: Filling empty places in the DV buffer.

The last DV usage method cannot be applied when a left frame is lost, since the left frame in the MVC structure only decoded with intra or inter blocks where the inter blocks are referenced to the past left frames.

4.2.3 COMBINED MOTION VECTOR AND DISPARITY VECTOR COPY

In a stereo video sequence, when a right frame is lost, the two former methods use either corresponding left motion vectors or last disparity vectors. However, in MVC, a P or B frame is encoded such that, intrapicture, interpicture and inter view predicted blocks may exist in the same frame. If only the left motion vectors are used, the intrapicture coded MBs existing in the left frame are assumed to have zero MV. Whereas corresponding MBs in the last right frame might be interview predicted. On the other hand, if last DVs are used to conceal a frame, then the intrapicture and interpicture coded MBs are assumed to have DV values obtained by histogramming the existing DVs. However the corresponding MBs in the left frame might be interpicture predicted. In this method the CMVC usage and DVC methods are combined to minimize the buffer entries which are filled with zero MVs or the DVs obtained by histogramming.

To conceal a frame, first a complementary-MV and a last-DV-MV buffer are constructed. When a left frame is decoded successfully, then the motion vectors are saved to the complementary-MV buffer exactly the same as in the complementary motion vector usage. When a right frame is decoded successfully, then to the last-DV-MV buffer, two kinds of information are saved for each block. First information is whether the block is compensated as DV or MV. Second is the vector value. When a right frame is lost, then the buffers are checked. If a complementary-MV exists, then the block is concealed as in CMVC method. If not, then the last-DV-MV buffer is checked. If the block is compensated with DV, then DVC method is used. If the block is compensated with MV in the last frame, then the MV is copied and the block is compensated with this MV and the last right frame. If none of the conditions are satisfied, then the block is concealed with FC. The block scheme is given below.

In this method, different from the previous method, the buffer size is increased to more then two times the older buffers.

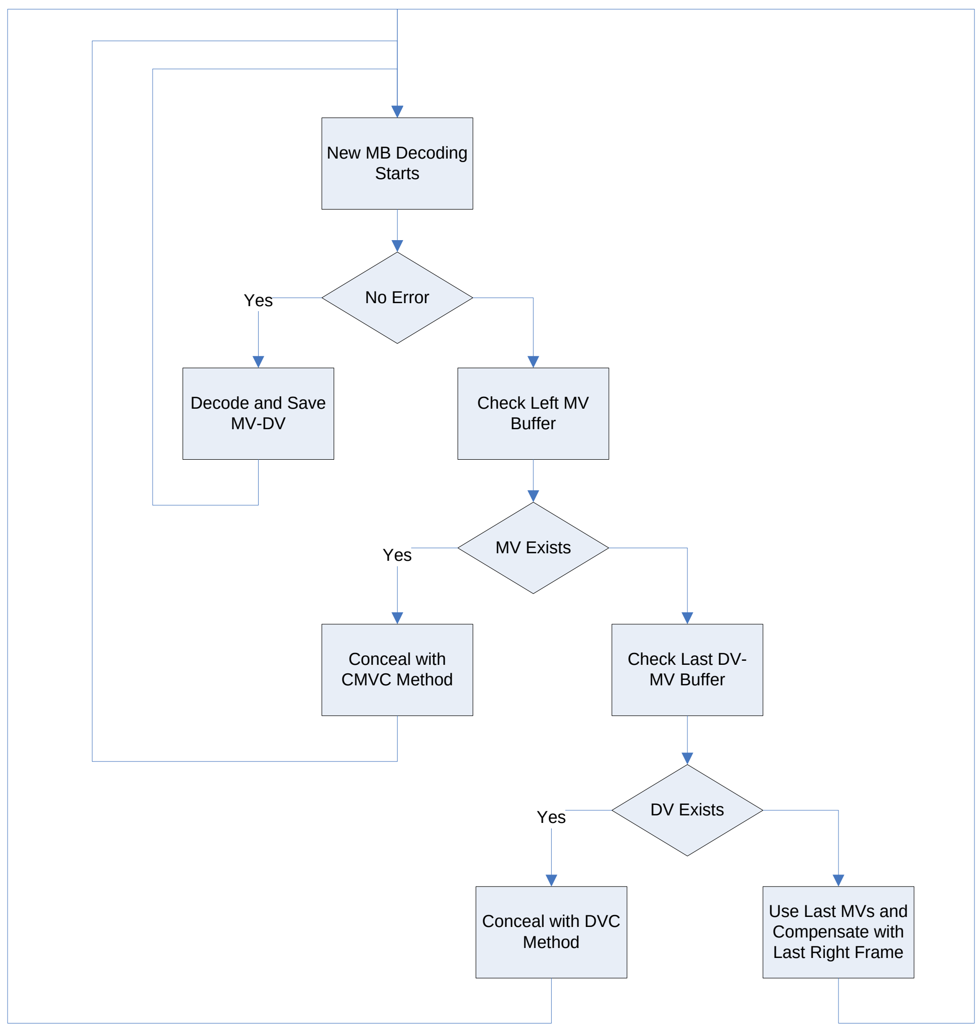


Figure 4-15: Flow chart for the CMVDVC method

CHAPTER 5

EXPERIMENTAL RESULTS

5.1 INTRODUCTION

The implemented methods are tested with four video sequences, which has the properties given below.

Table 5-1: The video sequences used in simulations.

Video	WidthxHeight	Video Characteristics	Screenshot
Heidelberg Alleys	432x240	Low Motion/High Detail	
Knights Quest	432x240	Animation	
Rhine Valley Moving	432x240	High Camera And Object Motion/Low Detail	
Rollerblade	320x240	High Object Motion/High Detail	

The Heidelberg video, which contains high detailed scenes of outdoor and city life, has been created by Dongleware [35]. The video has moderate and low motion characteristics and high depth complexity. The Knights Quest video has been produced by Red Star Studio Ltd. [36]. It is a computer animation which is an action scene, containing various types of object and camera movements. The Rhine Valley Moving video has been created by Cinovet [37]. In this video, low detailed nature, and transportation scenes are filmed. The video contains high camera movements and high object movements. The last video, Rollerblade, is created by FHG/HHI in cooperation with European Project “3D Phone” [38]. It contains the scenes from a competition, which has high object motions and almost zero camera movement.

The depth maps of the test sequences for the 2D+depth approach have been estimated using a hybrid recursive algorithm by FHG/HHI [39]. For the right view synthesis from the depth and left video sequences, the simple algorithm presented by Merke et. al was used [40].

In order to test the concealment methods, each frame of the sequences considered to be lost. The simulation flowchart is given below.

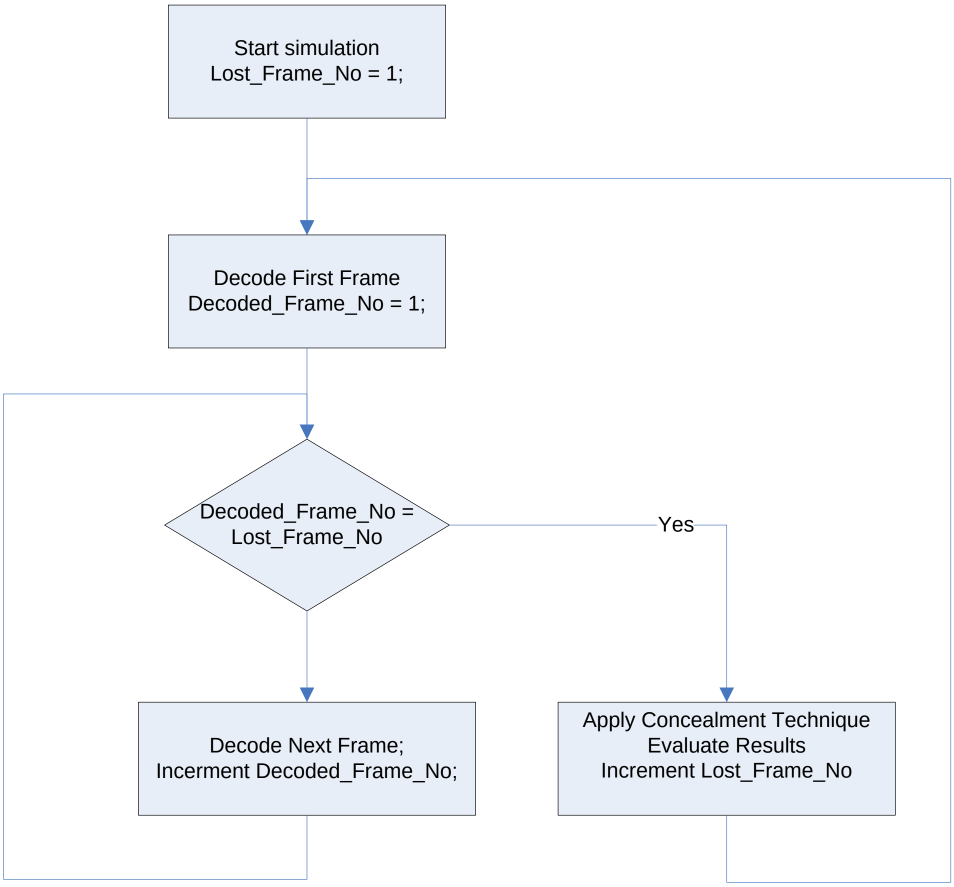


Figure 5-1: Simulation flowchart for the implemented methods.

With the simulation flowchart given above, it is ensured that for all of the frames of each of the sequences the methods are applied and results are taken.

Since the 3D error metric is a hot research topic, in this thesis more widely used two 2D error metrics are used. The performance evaluations are made by using the Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM). PSNR of an image with respect to the original image is calculated with:

$$\mathrm{PSNR} = 10 * \log\left(\frac{(255^2)*(\mathrm{width})*(\mathrm{height})}{\sum_{i=0}^{(\mathrm{height}-1)}\sum_{j=0}^{(\mathrm{width}-1)}(P(i,j)-P_{orig}(i,j))^2}\right) \quad (1)$$

where P is the image to be compared to the original image, and P_{orig} is the original image.

PSNR is a pixel-wise error metric which do not considers human eye perceptual characteristics. However SSIM is a measure of the two images which is more consistent with the human eye perception [41]. The formula of the SSIM model is given below:

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (2)$$

where x and y being two NxN sized window to be compared,

μ_x: average of x, μ_y: average of y,

σ_x^2: variance of x, σ_y^2: variance of y,

σ_{xy}: the covariance of x and y,

$c_1=(k_1L)^2$, $c_2=(k_2L)^2$:two variables to stabilize the division with weak denominator,

L: the dynamic range of pixel values,

$k_1 = 0.01$ and $k_2 = 0.03$ by default.

Default window is Gaussian given by;

window = fspecial('gaussian', 11, 1.5);

For the SSIM calculations the code at [42] is used with default configuration.

5.2 EXPERIMENTAL RESULTS

5.2.1 EVALUATIONS OF EC METHODS FOR 2D+Depth

In the first method, in case of a frame loss (left or depth) the right frames are concealed with either Frame Copy Before Rendering method or direct right frame copy. The horizontal axis is the concealed frame number. The PSNR and SSIM results are taken for only the concealed frame. Cumulative effect of the concealment is not considered.

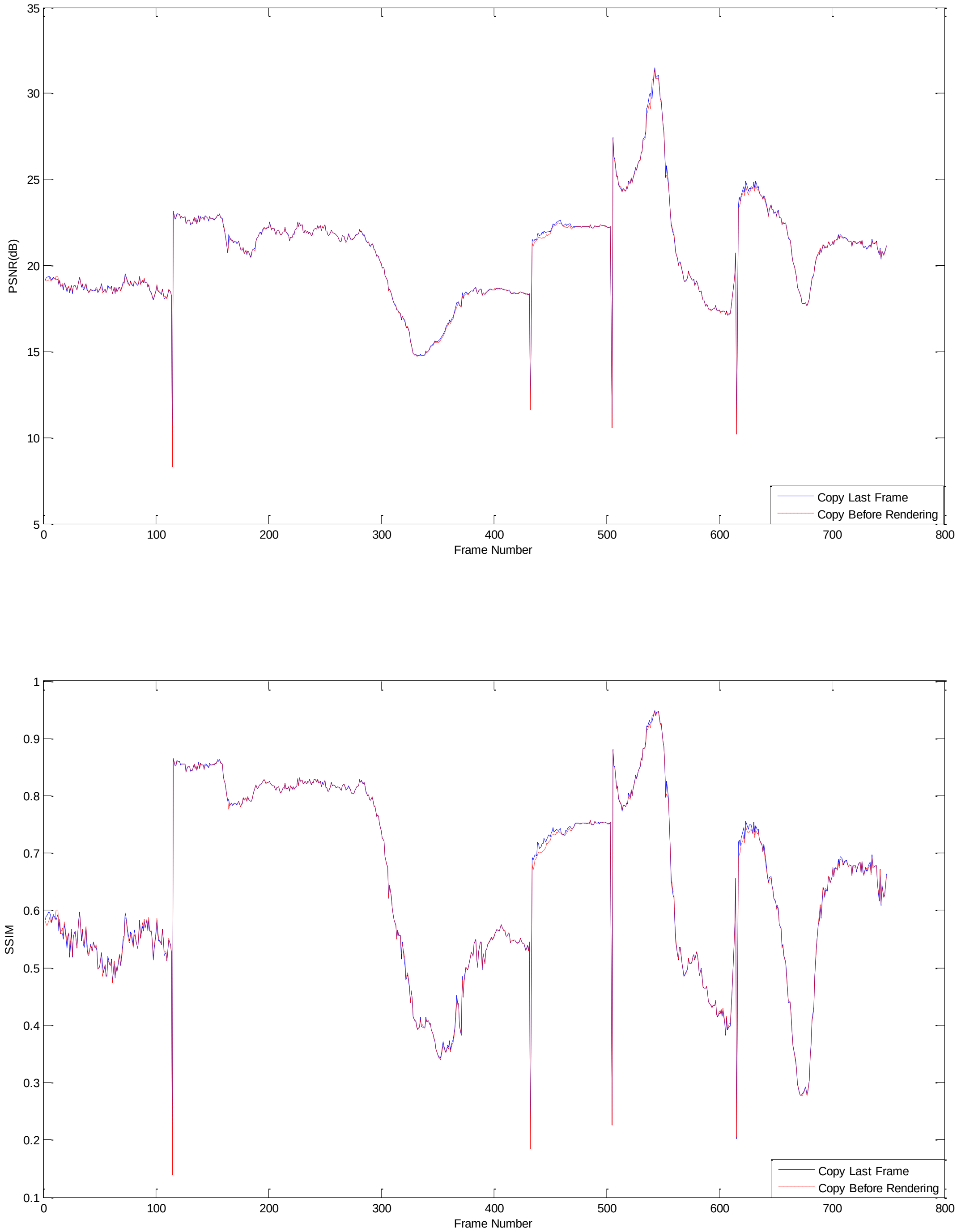


Figure 5-2: PSNR and SSIM results when left frames of "Heidelberg Alleys" video are lost and right frame is concealed with copying last right frame and copying last left frame before rendering.

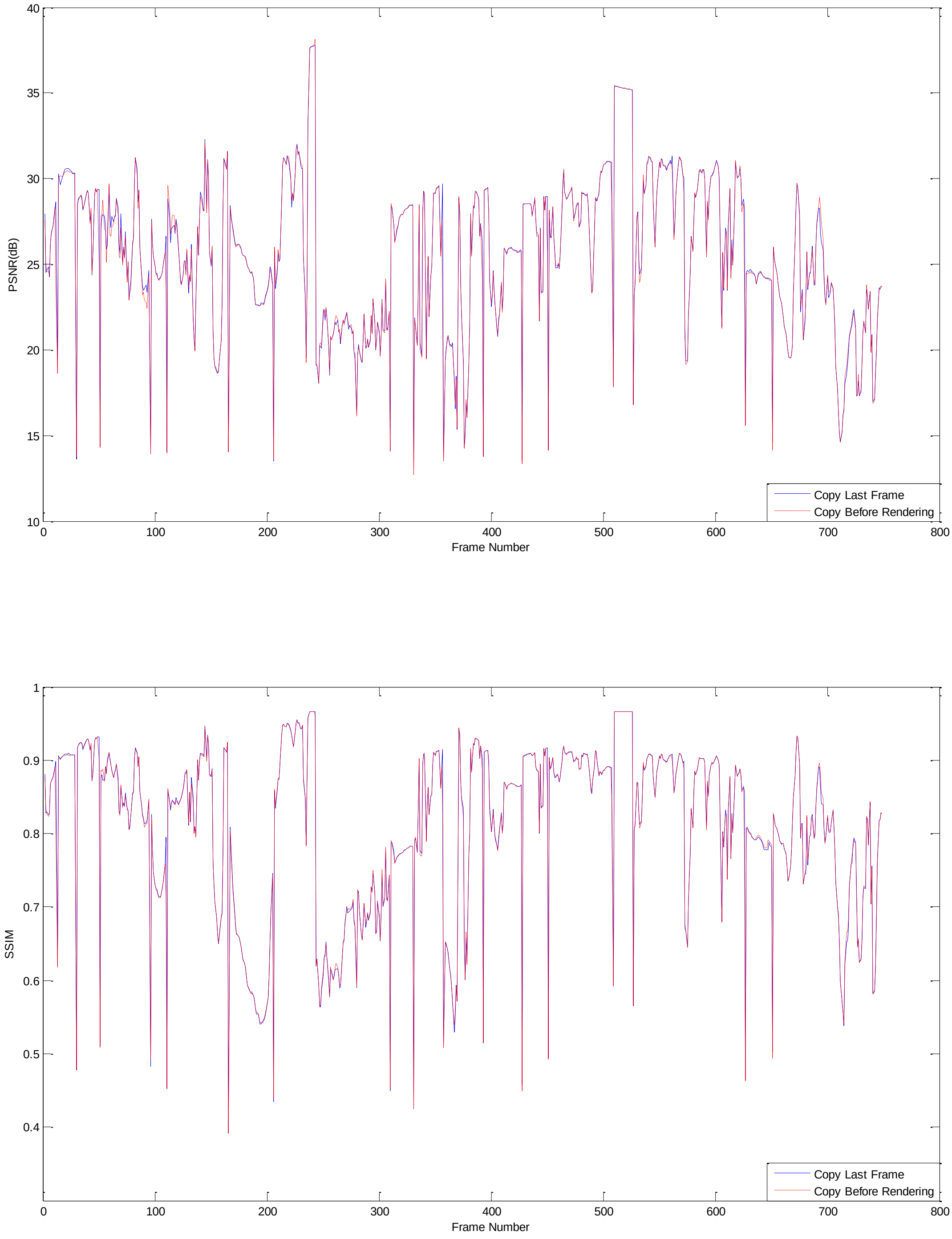


Figure 5-3: PSNR and SSIM results when left frames of "Knights Quest" video are lost and right frame is concealed with copying last right frame and copying last left frame before rendering.

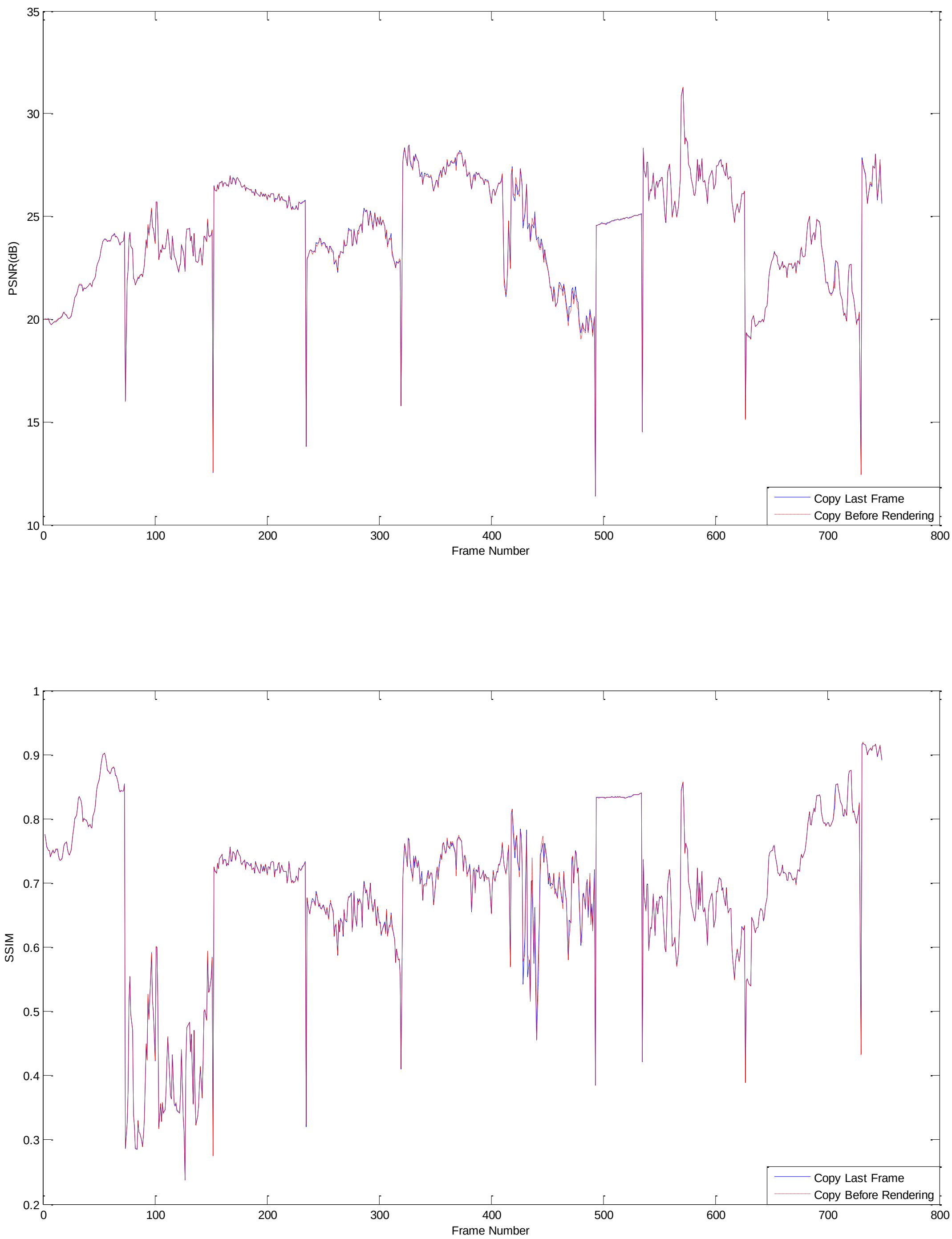
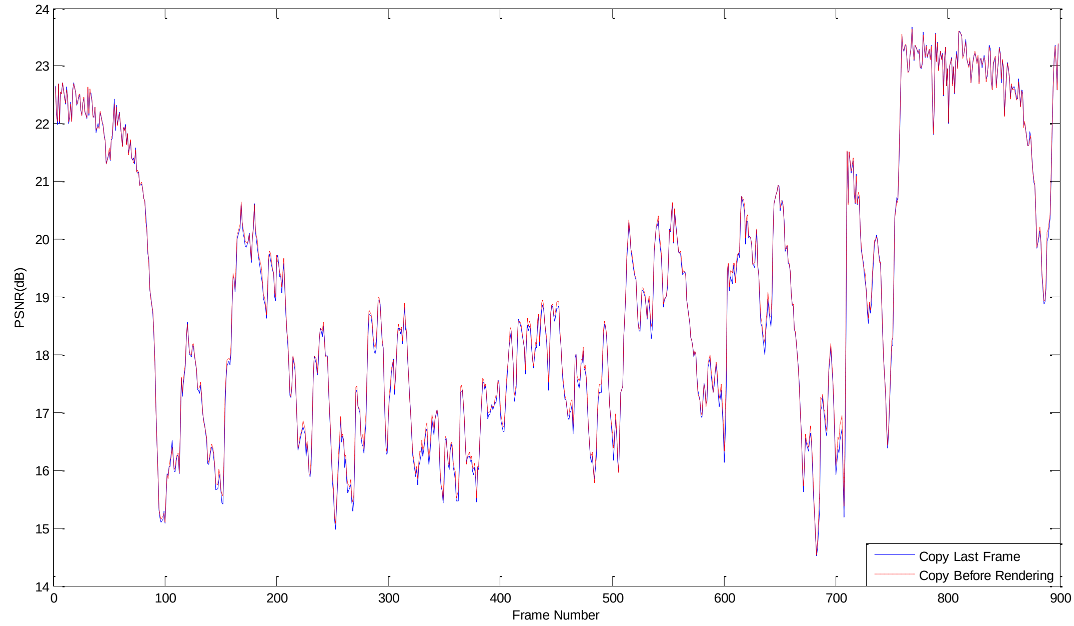


Figure 5-4: PSNR and SSIM results when left frames of "Rhine Valley Moving" video are lost and right frame is concealed with copying last right frame and copying last left frame before rendering.



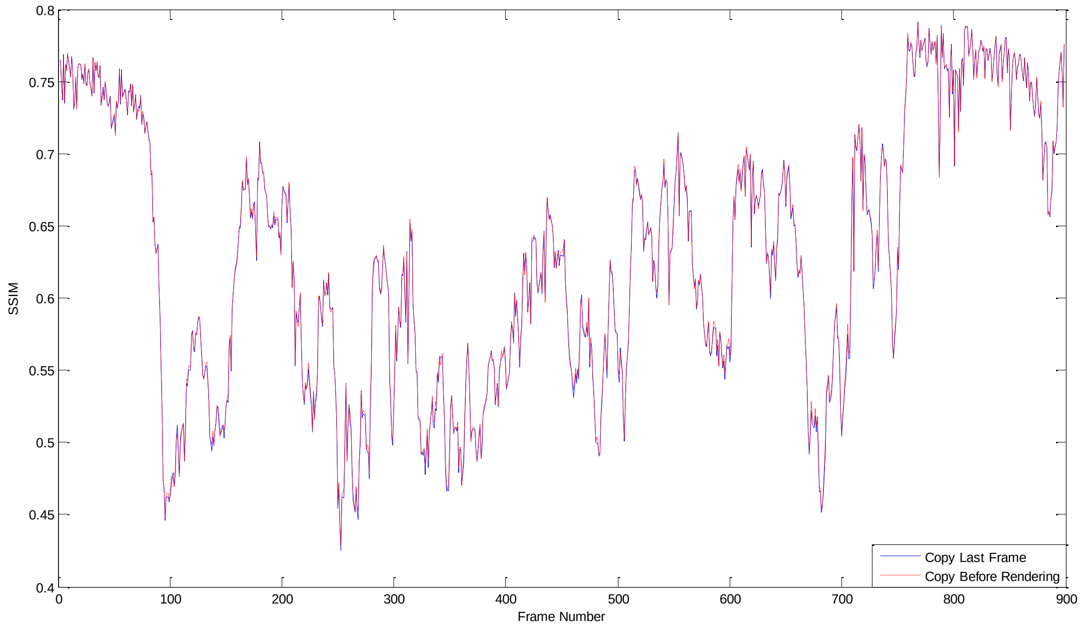


Figure 5-5: PSNR and SSIM results when left frames of "Rollerblade" video are lost and right frame is concealed with copying last right frame and copying last left frame before rendering.

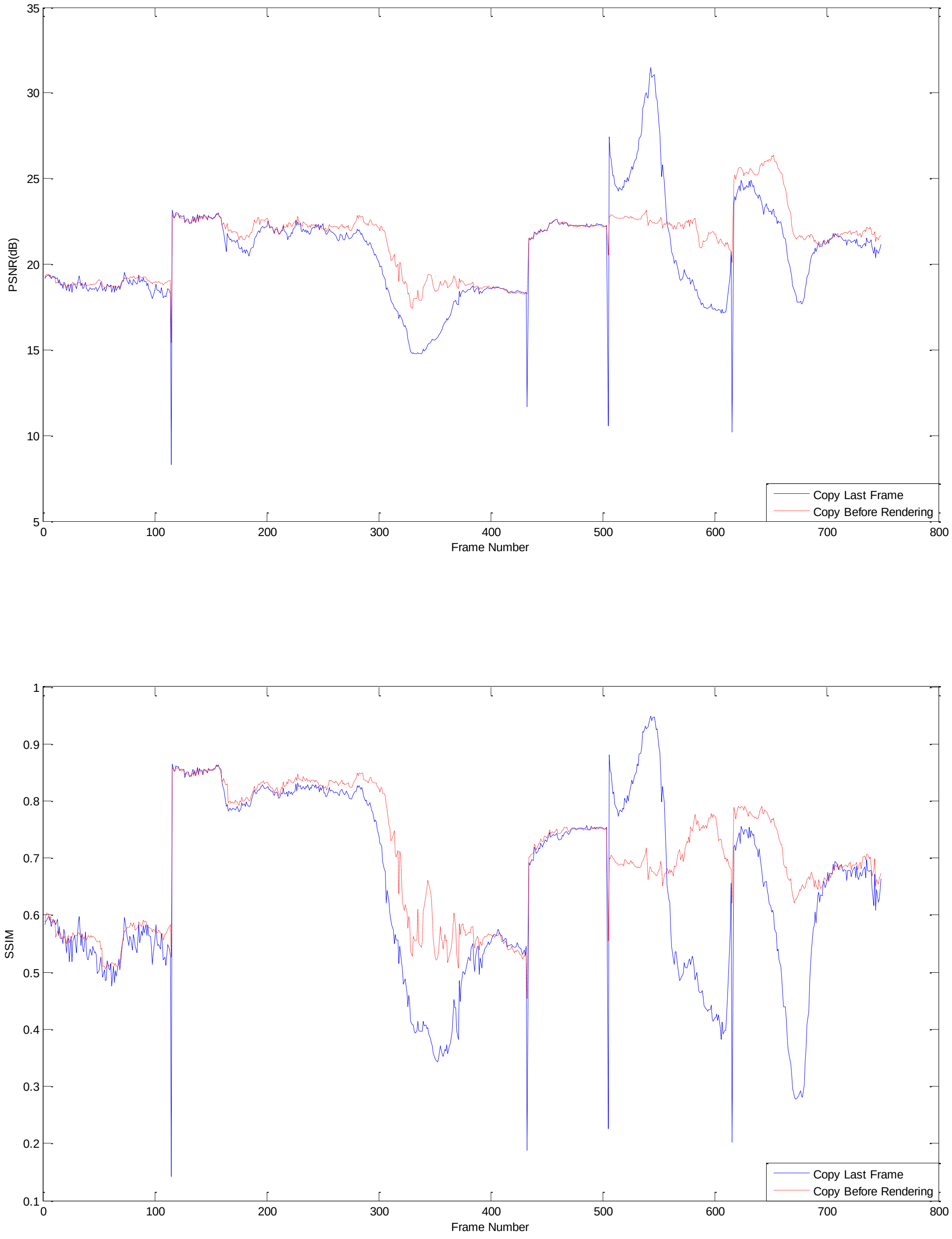
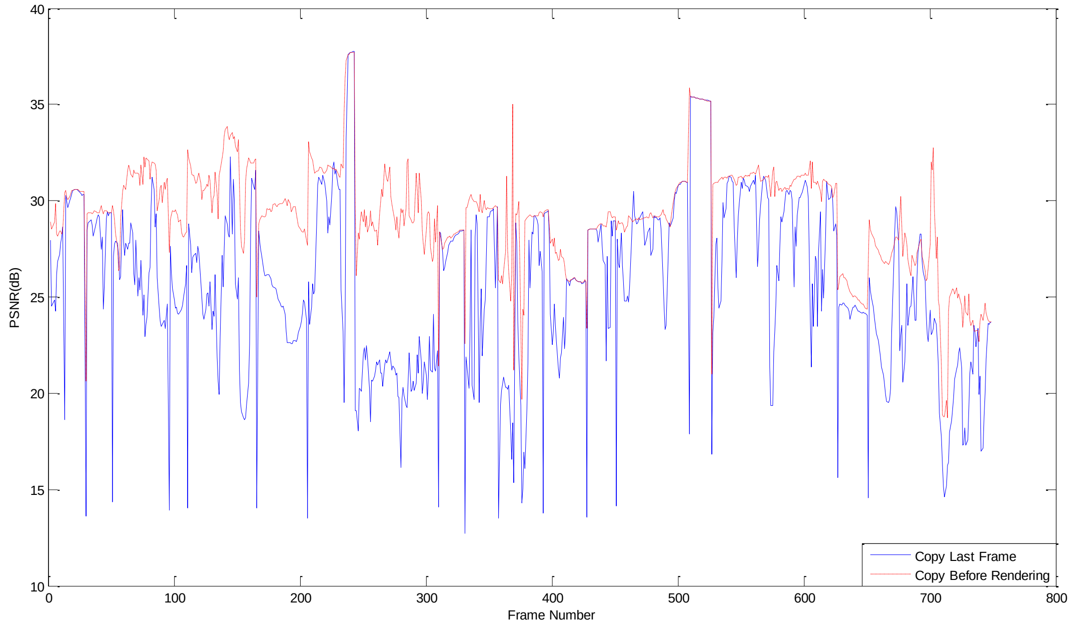


Figure 5-6: PSNR and SSIM results when depth frames of “Heidelberg Alleys” video are lost and right frame is concealed with copying last right frame and copying last depth frame before rendering.



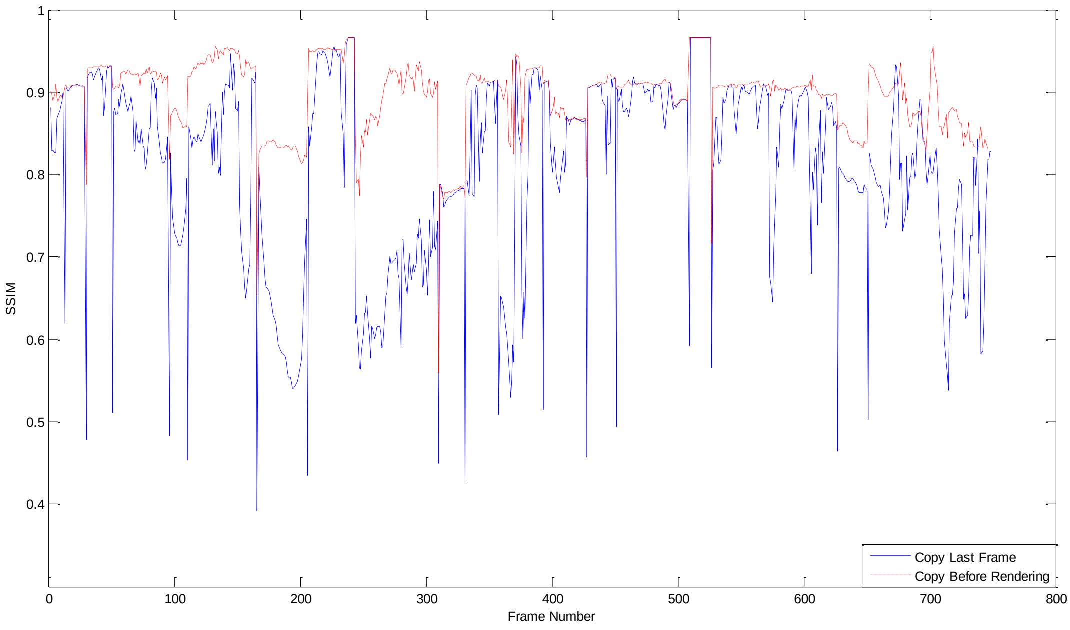
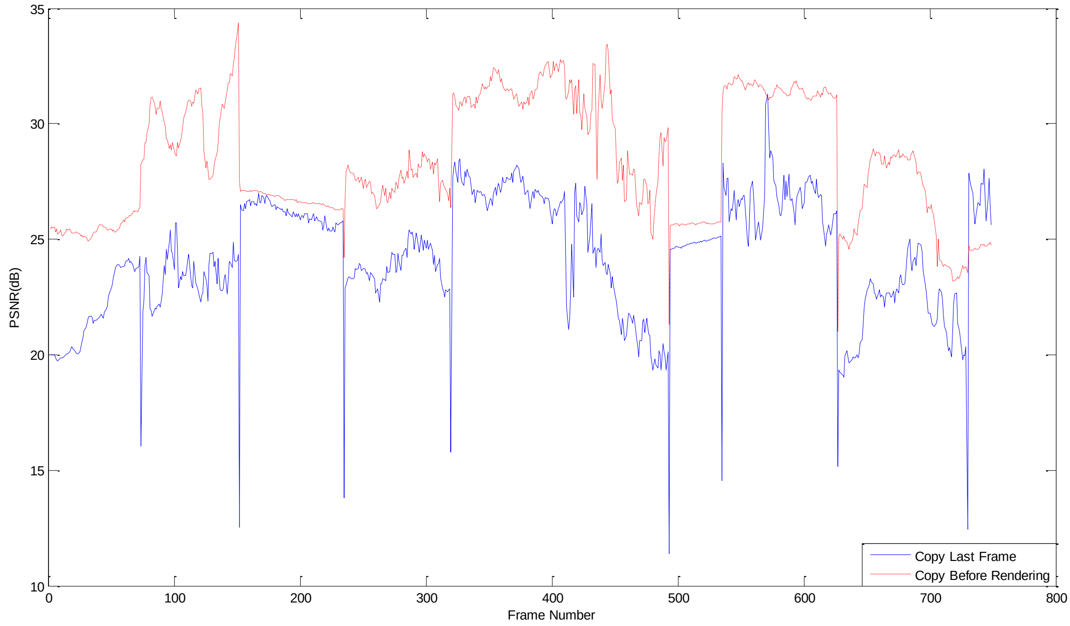


Figure 5-7: PSNR and SSIM results when depth frames of "Knights Quest" video are lost and right frame is concealed with copying last right frame and copying last depth frame before rendering.



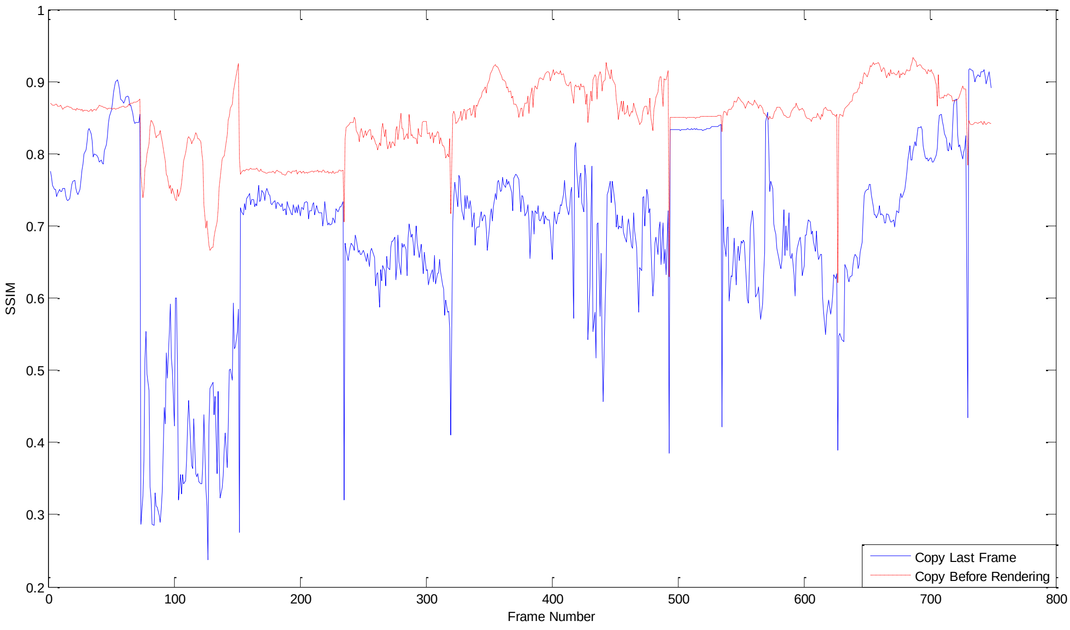
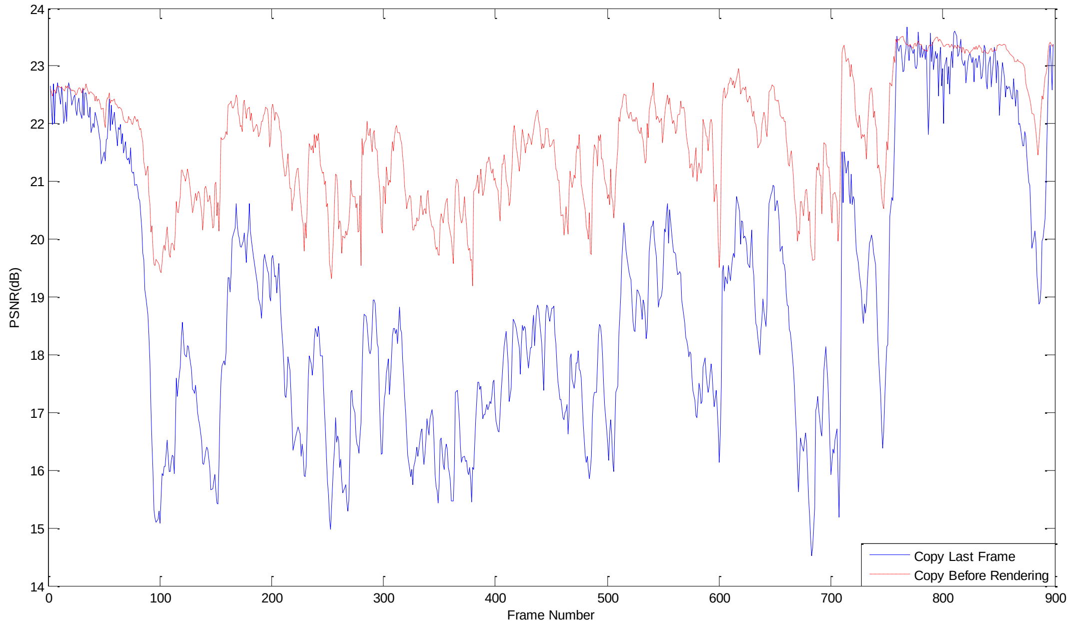


Figure 5-8: PSNR and SSIM results when depth frames of "Rhine Valley Moving" video are lost and right frame is concealed with copying last right frame and copying last depth frame before rendering.



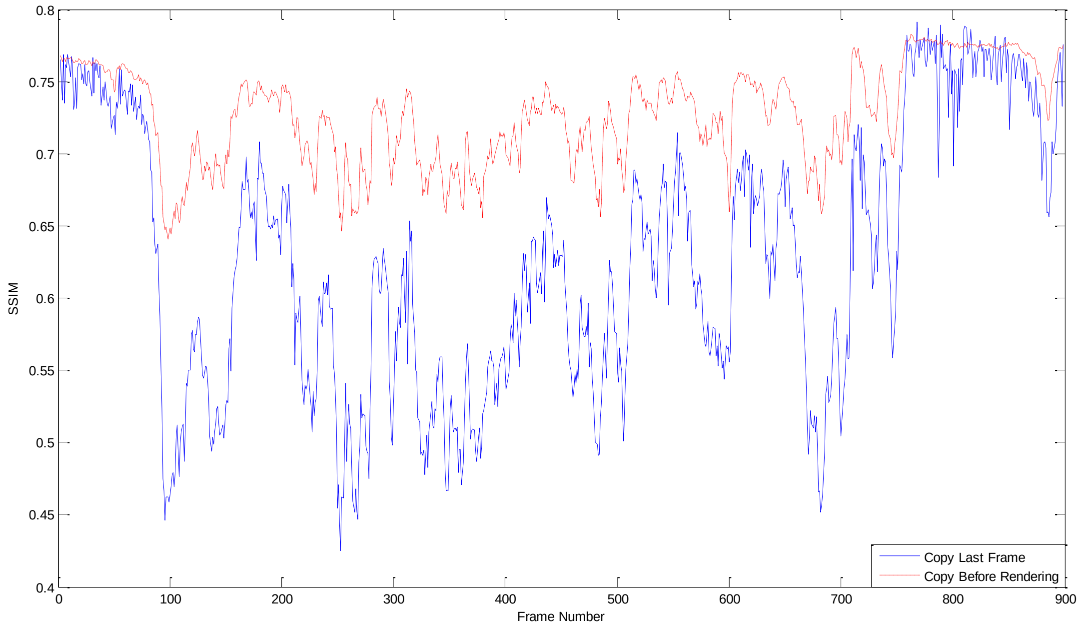


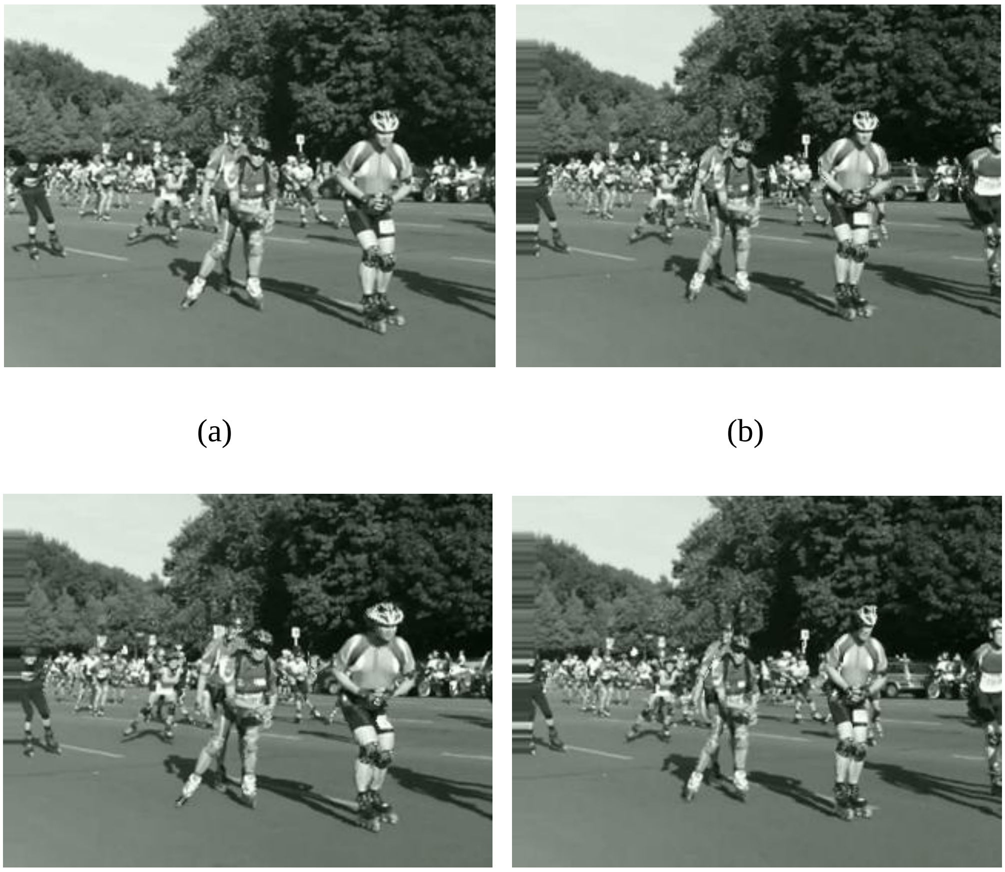
Figure 5-9: PSNR and SSIM results when depth frames of "Rollerblade" video are lost and right frame is concealed with copying last right frame and copying last depth frame before rendering.

As can be seen from the Figure 5-2 to Figure 5-5 , Frame Copy Before Rendering method does not perform better than direct copy when the left frame is lost. The performances of the two methods are nearly the same. However, when the depth frame is lost, the Frame Copy Before Rendering method gives better results as given in the Figure 5-6 to Figure 5-9. As can be seen from the Table 5-2, average PSNR increases 4 dB and SSIM increases 0,15 at most. Examining the videos individually, it can be seen that the FCBR method gives better results for moving scenes and last frame copy gives better results for still scenes than moving scenes. Between 510th and 615th frame, the scene in the Heidelberg video is a simple camera movement. Up to 560th frame, PSNR and SSIM results of last right frame copy are higher. After 560th frame, the change in the viewing angle speeds up, so performance of last right frame copy diminishes. However, the performance of FCBR between the 510th and 615th frames does not change too much, and better than frame copy in the 560-615 interval. In the graphs, the notches are the scene changes where the former frame is not correlated with the succeeding frames.

Table 5-2: Average PSNR and SSIM results for the Frame Copy and Frame Copy Before Rendering methods.

Video	Lost View	PSNR		SSIM	
		Frame Copy	Frame Copy Before Rendering	Frame Copy	Frame Copy Before Rendering
Heidelberg Alleys	Left	20,66	20,63	0,6483	0,6474
Knights Quest		25,60	25,55	0,8085	0,8079
Rhine Valley Moving		24,24	24,22	0,6911	0,6909
Rollerblade		19,00	19,04	0,6281	0,6290
Heidelberg Alleys	Depth	20,66	21,31	0,6483	0,6983
Knights Quest		25,60	29,27	0,8085	0,8962
Rhine Valley Moving		24,24	28,33	0,6911	0,8483
Rollerblade		19,00	21,70	0,6281	0,7271

The reason, why the FCBR method gives better results for depth losses is because the information carried in the depth frame is smaller than the left frame. When a depth frame is lost, instead of using last right frame, even copying the left frame to the place of right frame and switching from 3D to 2D is preferable as given in [29] because the temporal information is not lost for that frame. As for FCBR method, using current left frame and last depth frame does not lose the 3D perception too much and gives good PSNR and SSIM results compared to copying last frame. For both depth frame and left frame losses, there are no blocking effects but artifacts, as can be seen in Figure 5-10.

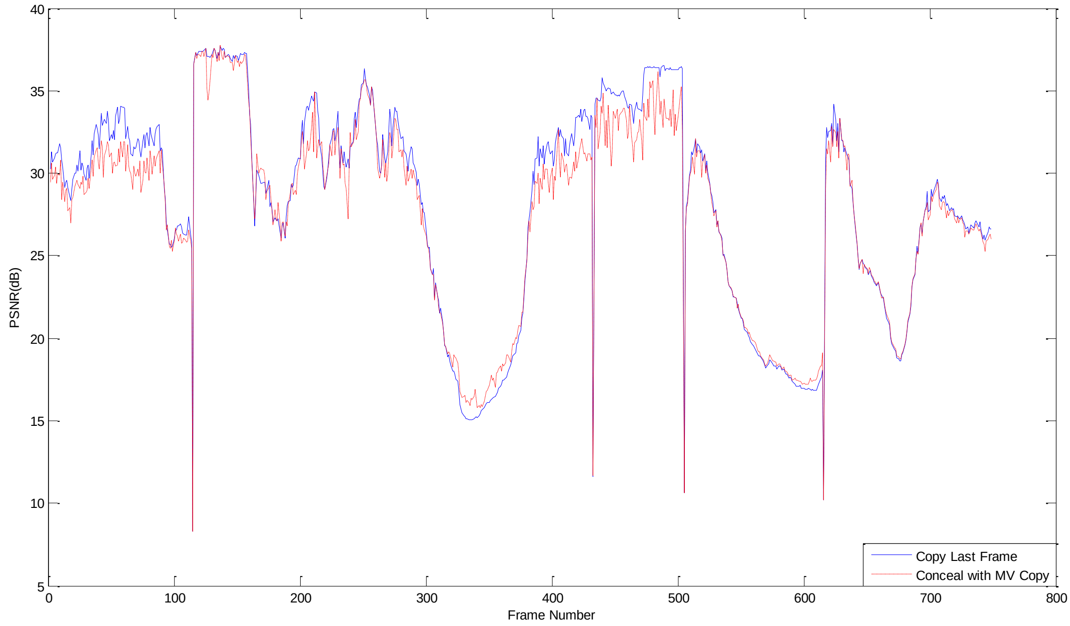


(a) (b)

(c) (d)

Figure 5-10: (a) 262nd right frame of original Rollerblade video, (b) 261st right frame of rendered Rollerblade video, (c) 262nd right frame of Rollerblade video is concealed with FCBR when depth frame is lost (d) 262nd right frame of Rollerblade video is concealed with FCBR when left frame is lost

In the second method, Complementary Motion Vector Usage for 2D+depth is investigated where instead of right image; the lost image itself is concealed. Then the right frame can be rendered from the concealed frames. If a left frame is lost, it is concealed with the information provided with depth sequence, and vice versa i.e. the MVs of the successfully received complementary view are copied. The horizontal axis is the concealed frame number in the graphs. The PSNR and SSIM results are taken for only the concealed frame with respect to the original frame. Cumulative effect of the concealment is not considered.



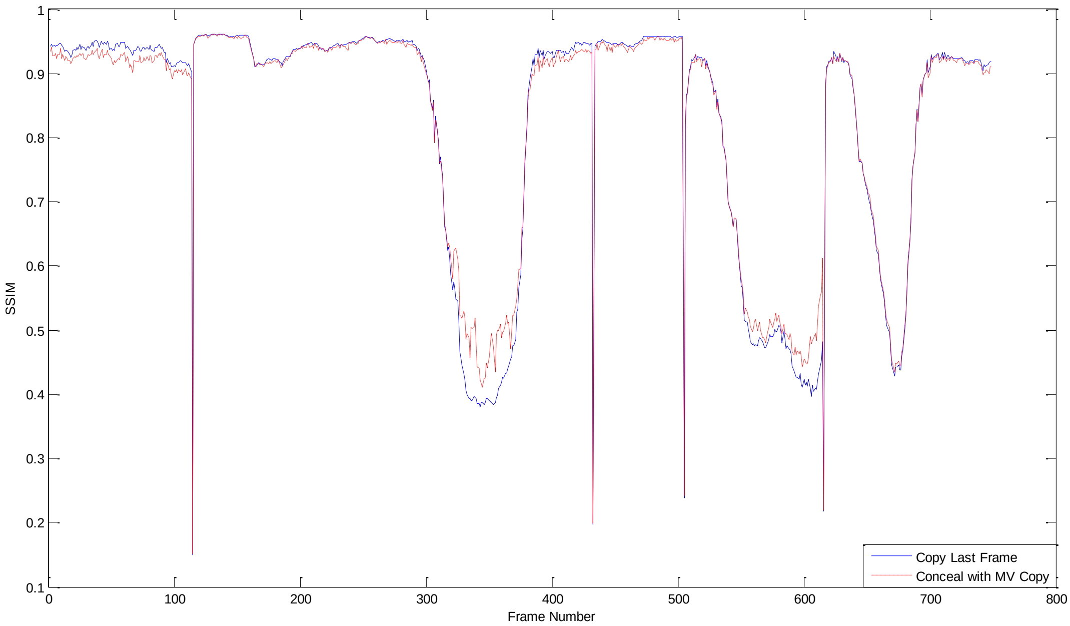
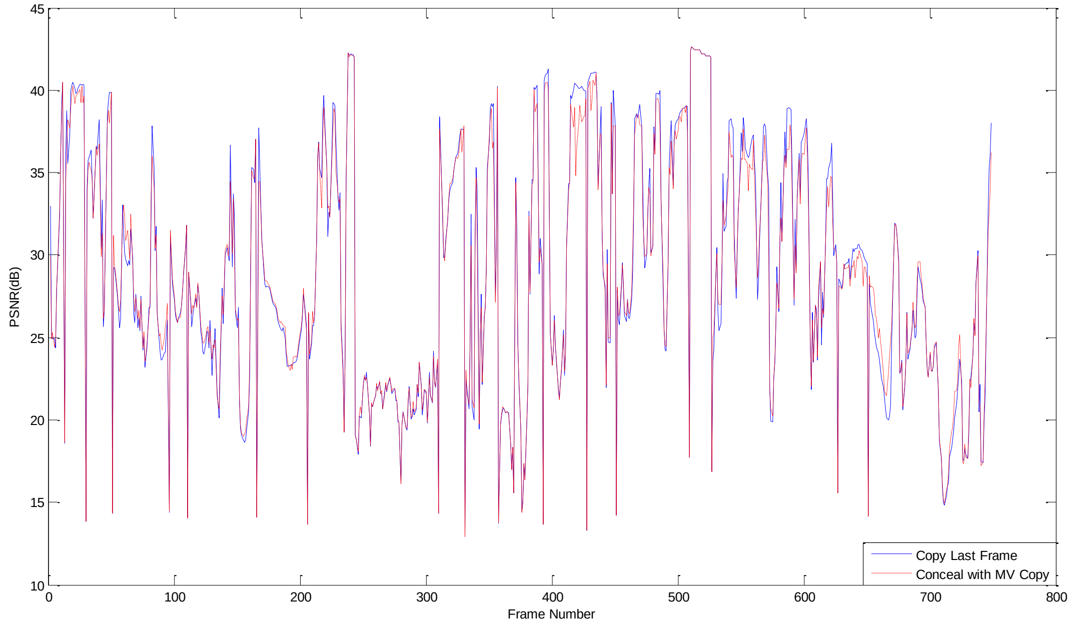


Figure 5-11: PSNR and SSIM results when left frames of "Heidelberg Alleys" video are lost and concealed with copying last left frame and copying depth MVs.



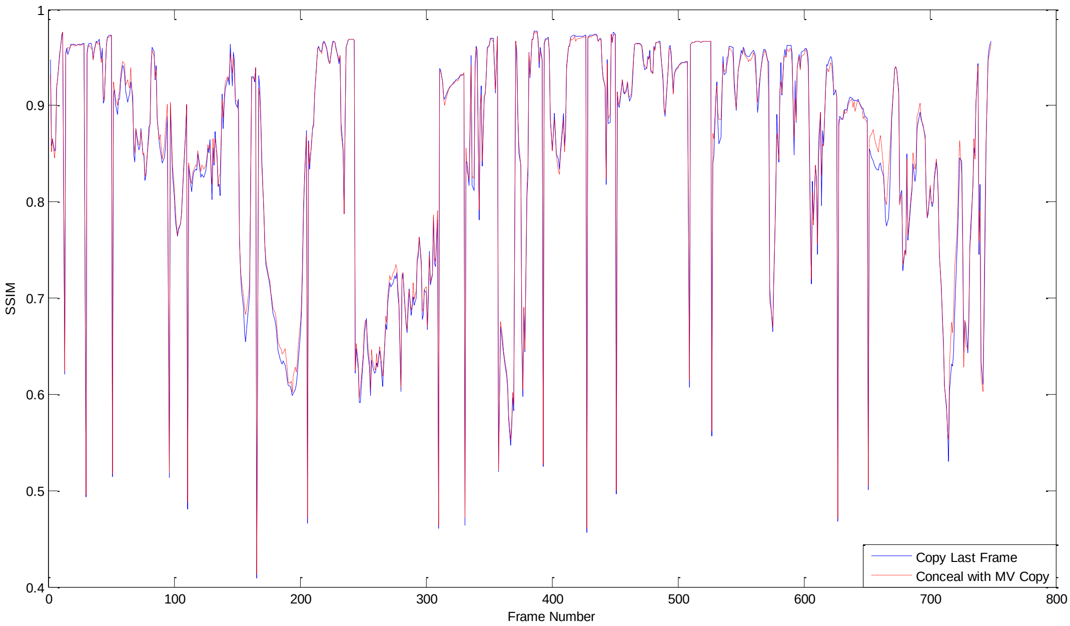


Figure 5-12: PSNR and SSIM results when left frames of "Knights Quest" video are lost and concealed with copying last left frame and copying depth MVs.

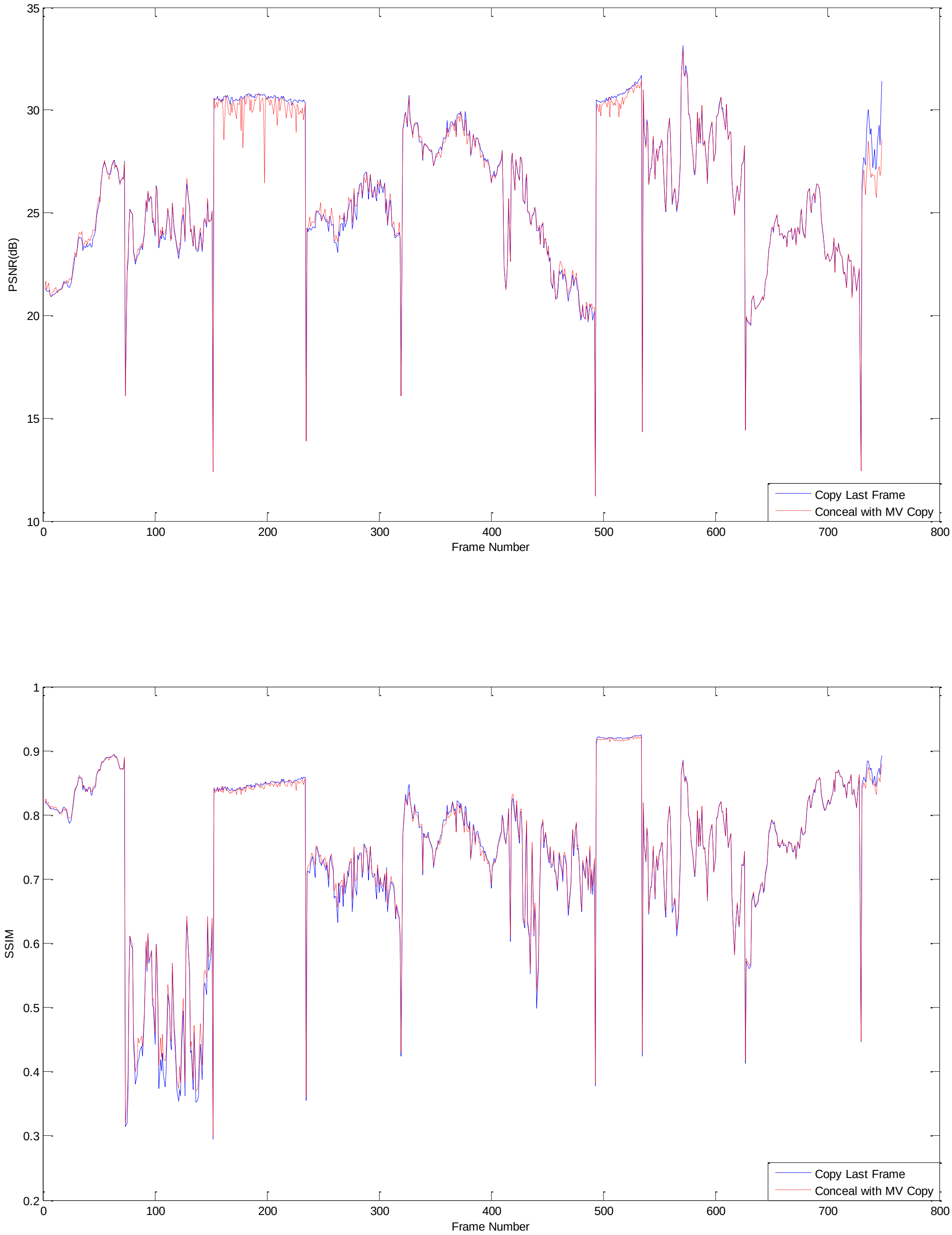
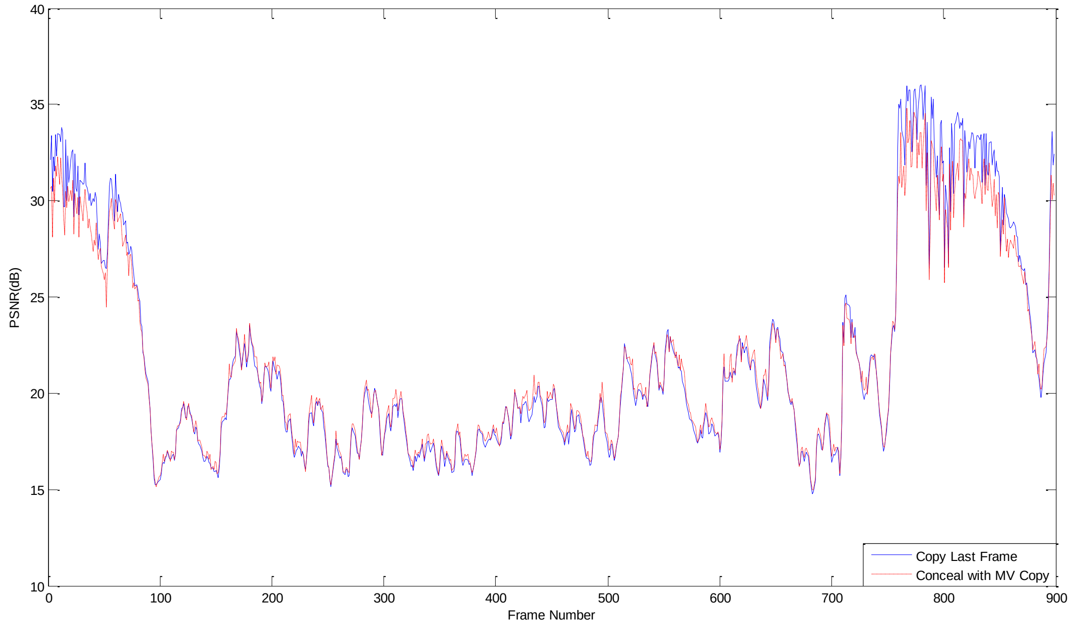


Figure 5-13: PSNR and SSIM results when left frames of “Rhine Valley Moving” video are lost and concealed with copying last left frame and copying depth MVs.



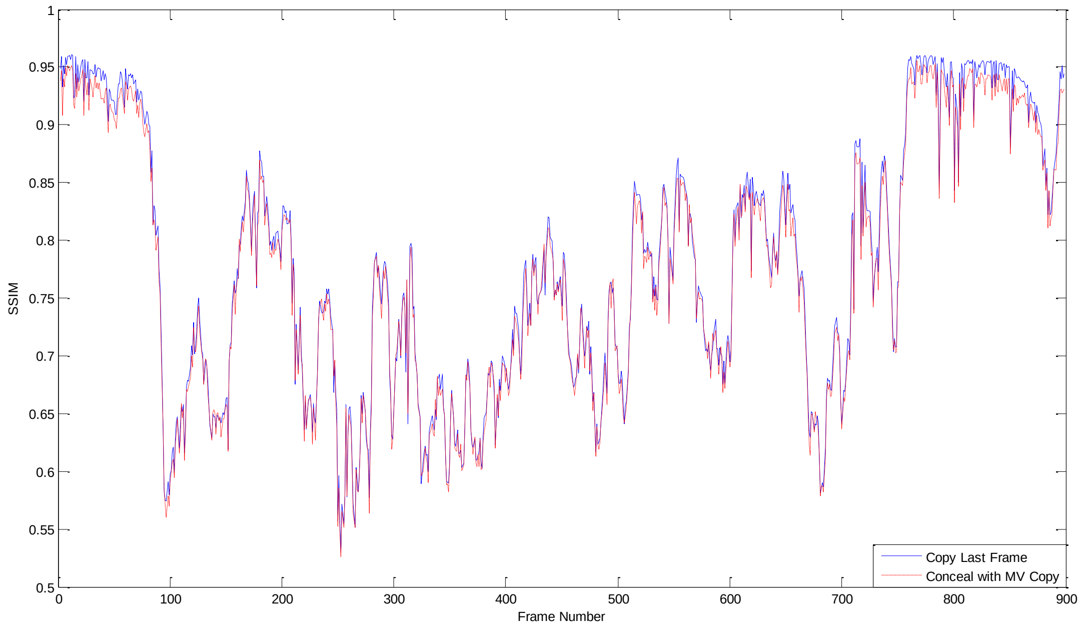


Figure 5-14: PSNR and SSIM results when left frames of "Rollerblade" video are lost and concealed with copying last left frame and copying depth MVs.

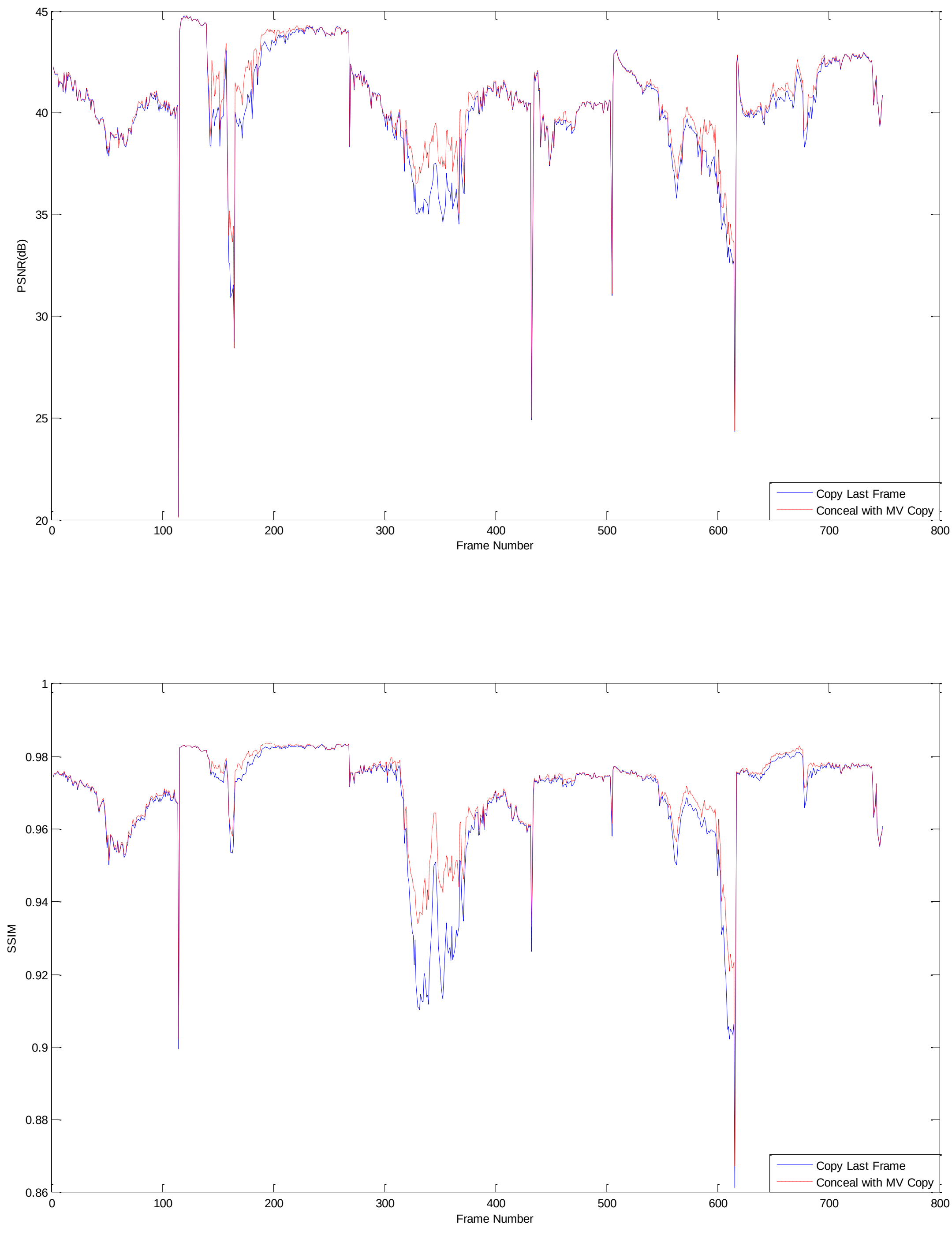


Figure 5-15: PSNR and SSIM results when depth frames of "Heidelberg Alleys" video are lost and concealed with copying last depth frame and copying left MVs.

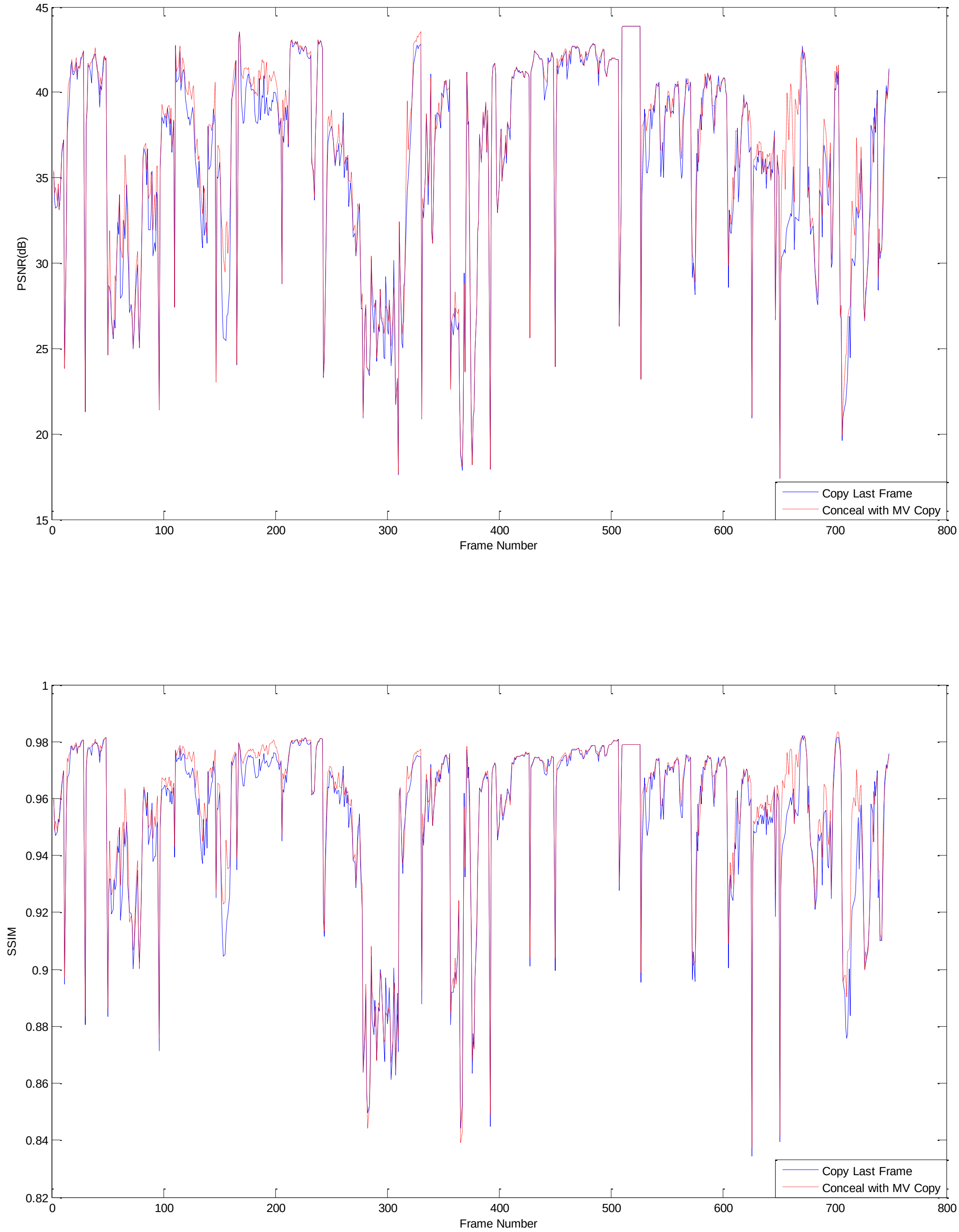
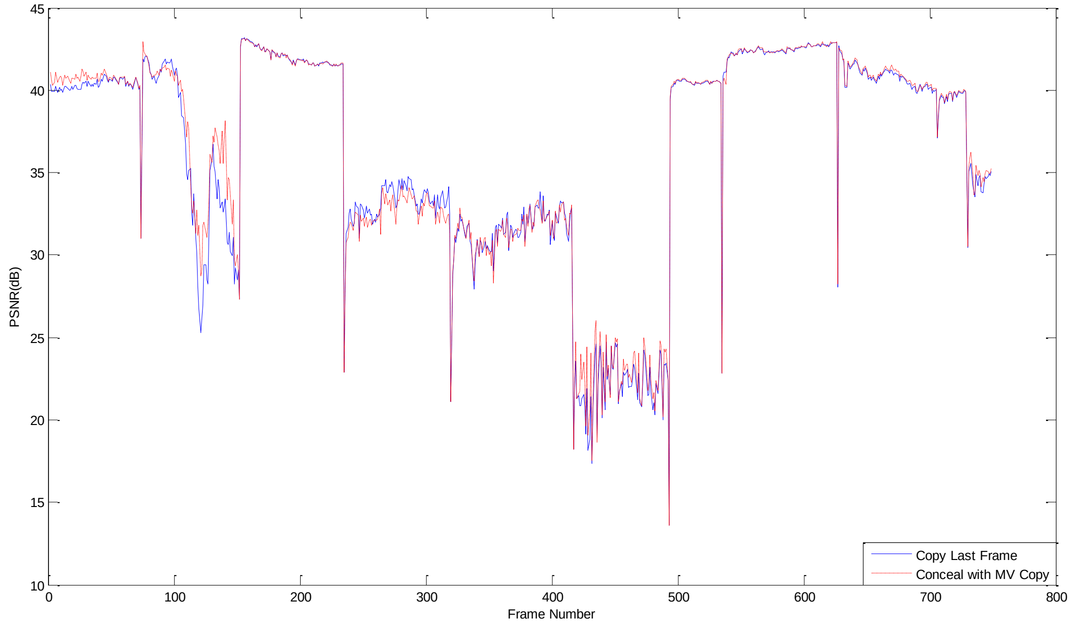


Figure 5-16: PSNR and SSIM results when depth frames of “Knights Quest” video are lost and concealed with copying last depth frame and copying left MVs.



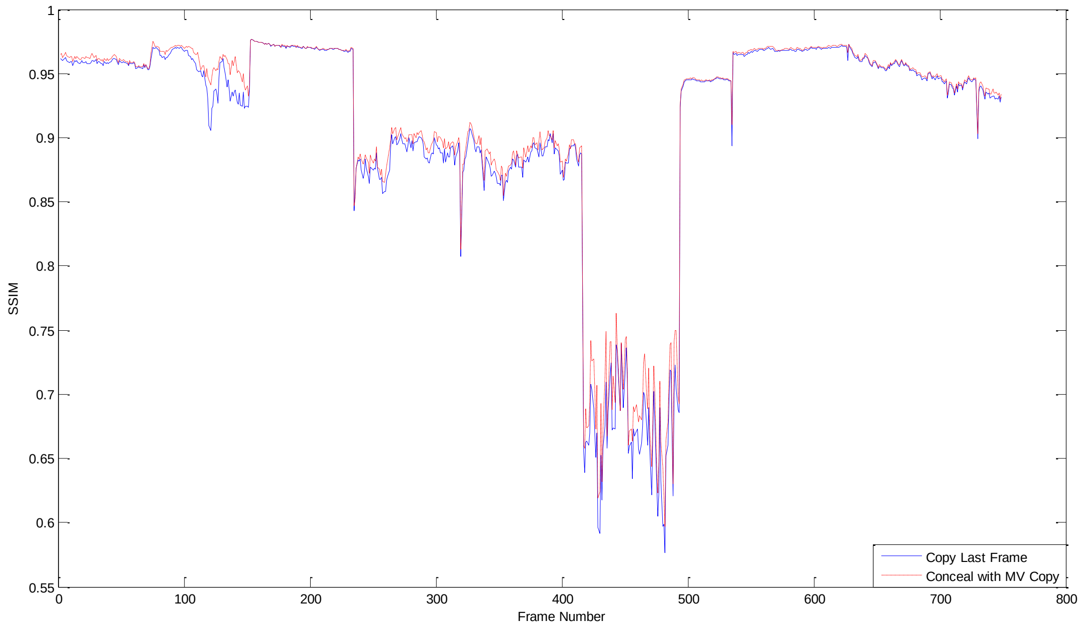


Figure 5-17: PSNR and SSIM results when depth frames of "Rhine Valley Moving" video are lost and concealed with copying last depth frame and copying left MVs.

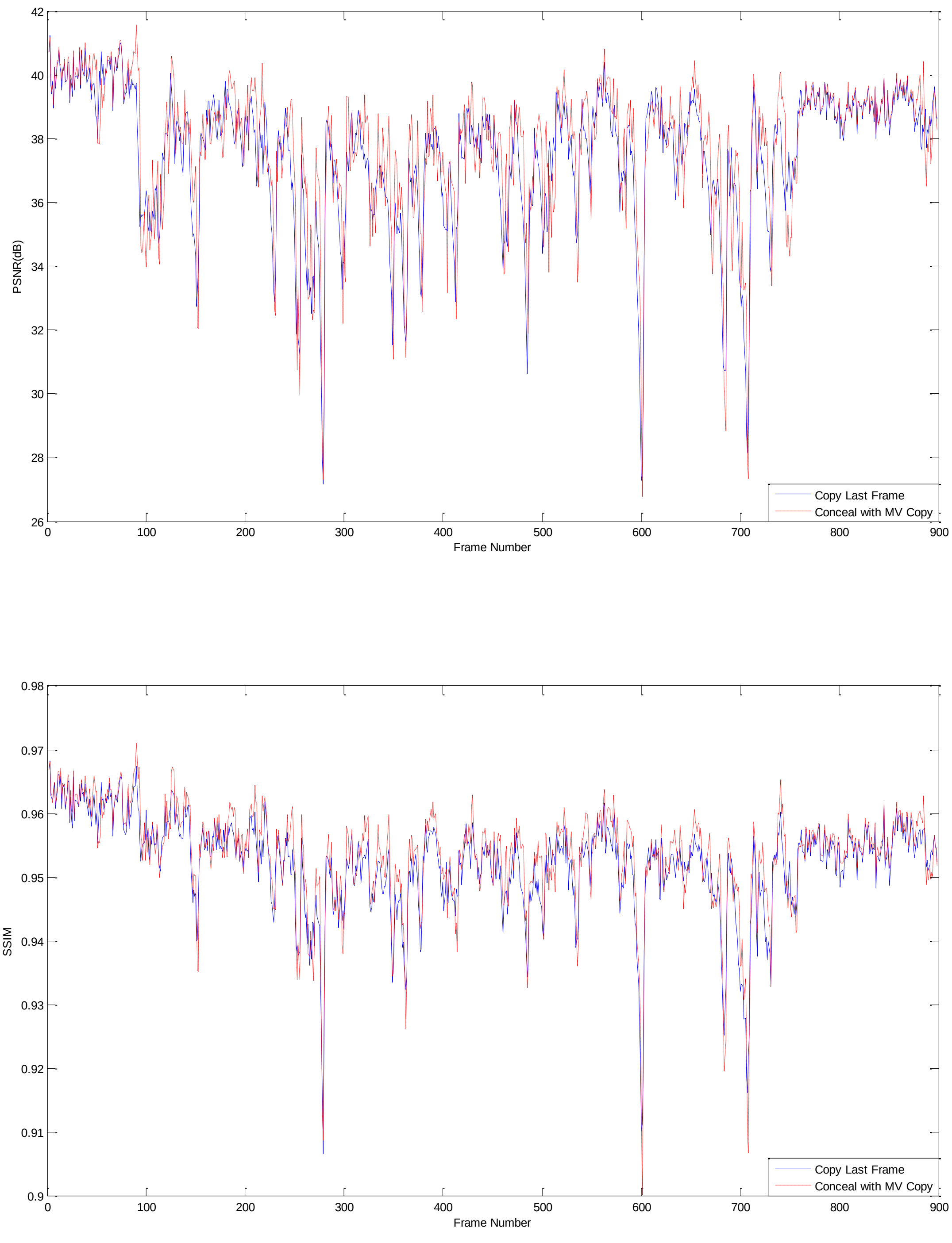


Figure 5-18: PSNR and SSIM results when depth frames of “Rollerblade” video are lost and concealed with copying last depth frame and copying left MVs.

The CMVC method does not improve the video quality in cases of left or depth frame losses. Only when the scene includes camera panning, the method gives better results. However it is not decisive for all cases. For instance in the Heidelberg video, those kind of scenes are concealed well, but in the Rhine Valley video, the method is not successful. Since the camera focus in the Rhine Valley video is distant from the camera, the depth video does not include textured information. Smooth image results with erroneous or zero MVs which cannot be used for the left frame concealment. In other cases, performance of the FC and CMVC methods are similar.

Table 5-3: Average PSNR and SSIM results for the Frame Copy and Complementary MV Copy Methods

Video	Lost View	PSNR		SSIM	
		Frame Copy	MV Copy	Frame Copy	MV Copy
Heidelberg Alleys	Left	28,15	27,58	0,8213	0,8251
Knights Quest		28,89	28,78	0,8491	0,8514
Rhine Valley Moving		26,06	26,00	0,7462	0,7485
Rollerblade		21,74	21,56	0,7800	0,7733
Heidelberg Alleys	Depth	40,32	40,72	0,9683	0,9709
Knights Quest		35,97	36,61	0,9537	0,9566
Rhine Valley Moving		36,24	36,43	0,9098	0,9153
Rollerblade		37,44	37,78	0,9521	0,9536

In the following pictures, in the frame which is concealed with CMVC method, there are artificial blocks. The places which are not visible to the sight and being visible in the next frame are problematic. When an object moves, the place behind it comes in to the vision, and in correctly received pictures this area is mostly coded as intrapicture predictive blocks. However, in lost frames, there is no information about the occluded area. In the CMVC method, that area is simply copied from the former frame if there is no motion information coming from the complementary frame. This causes the moving object to replicate. In following pictures, it can be seen that the concealed frame resembles to the last frame. Very few quantities of blocks are compensated with nonzero MVs, which is the result of the smooth or mostly intra coded depth frames. To obtain a natural scene, using last frame can be preferred. Also the blocks that are concealed with the CMVC method are not in the exactly correct places in the image because of erroneous MVs. Since the depth video is coarse and MV estimation may not be accurate with respect to a textured video, depth motion vector usage is not feasible for the left video concealment.

(a)

(b)

(c)

Figure 5-19: (a) 262nd left frame of original Rollerblade video, (b) 261st left frame of decoded Rollerblade video, (c) 262nd left frame of Rollerblade video is concealed with CMVC

5.2.2 EVALUATIONS OF EC METHODS FOR STEREO VIDEO

In this part the three stereo error concealment techniques will be evaluated. In the first method the effect of only copying complementary MVs is implemented, whereas in the second method the effect of copying last DVs is implemented. In the last method, combination of copying complementary MVs and last DVs-MVs will be investigated. The second and third methods are tested for only right frame losses, because the left frame is only compensated with MVs where last left frames are the references for compensation.

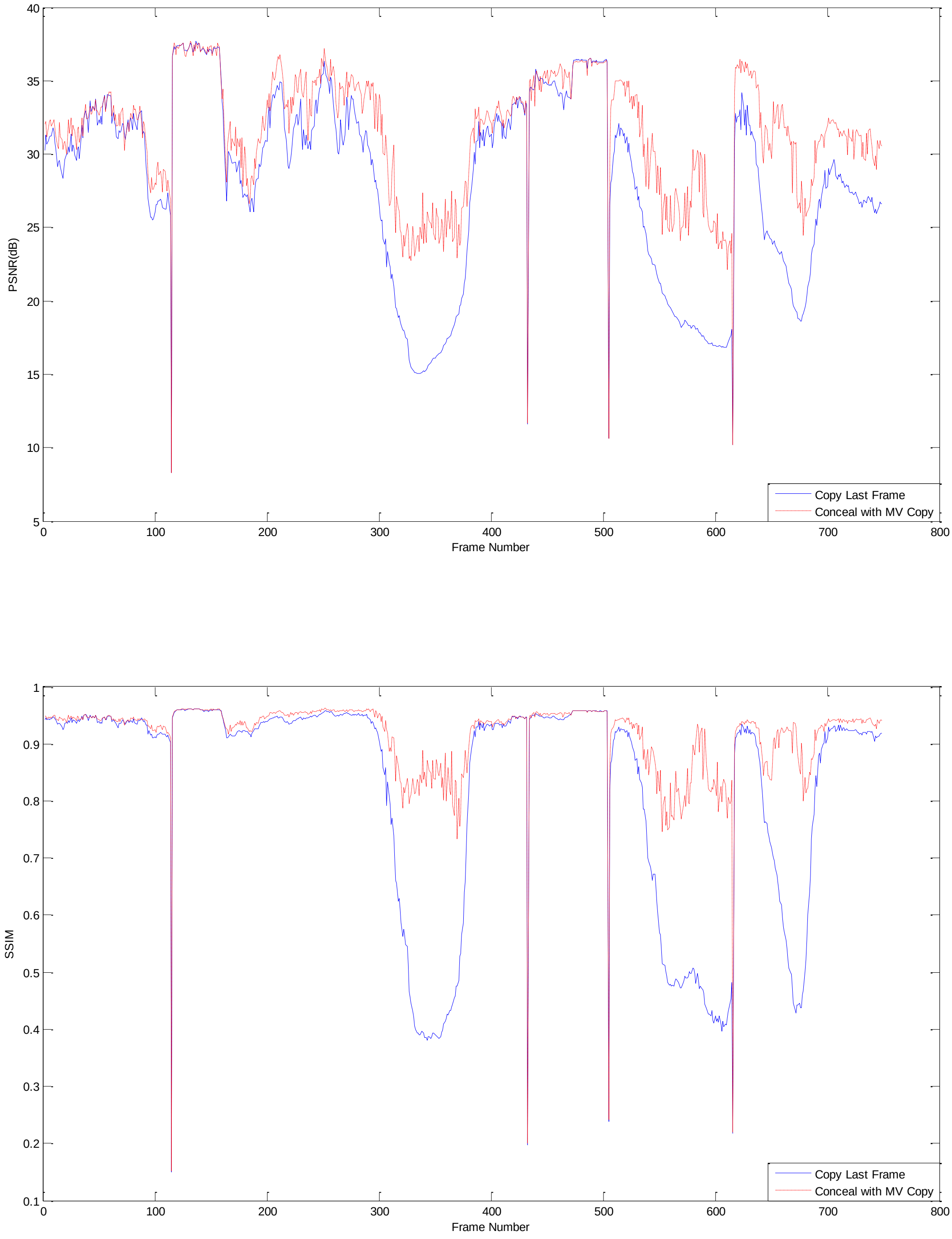


Figure 5-20: PSNR and SSIM results when left frames of "Heidelberg Alleys" video are lost and concealed with copying last left frame and copying right MVs.

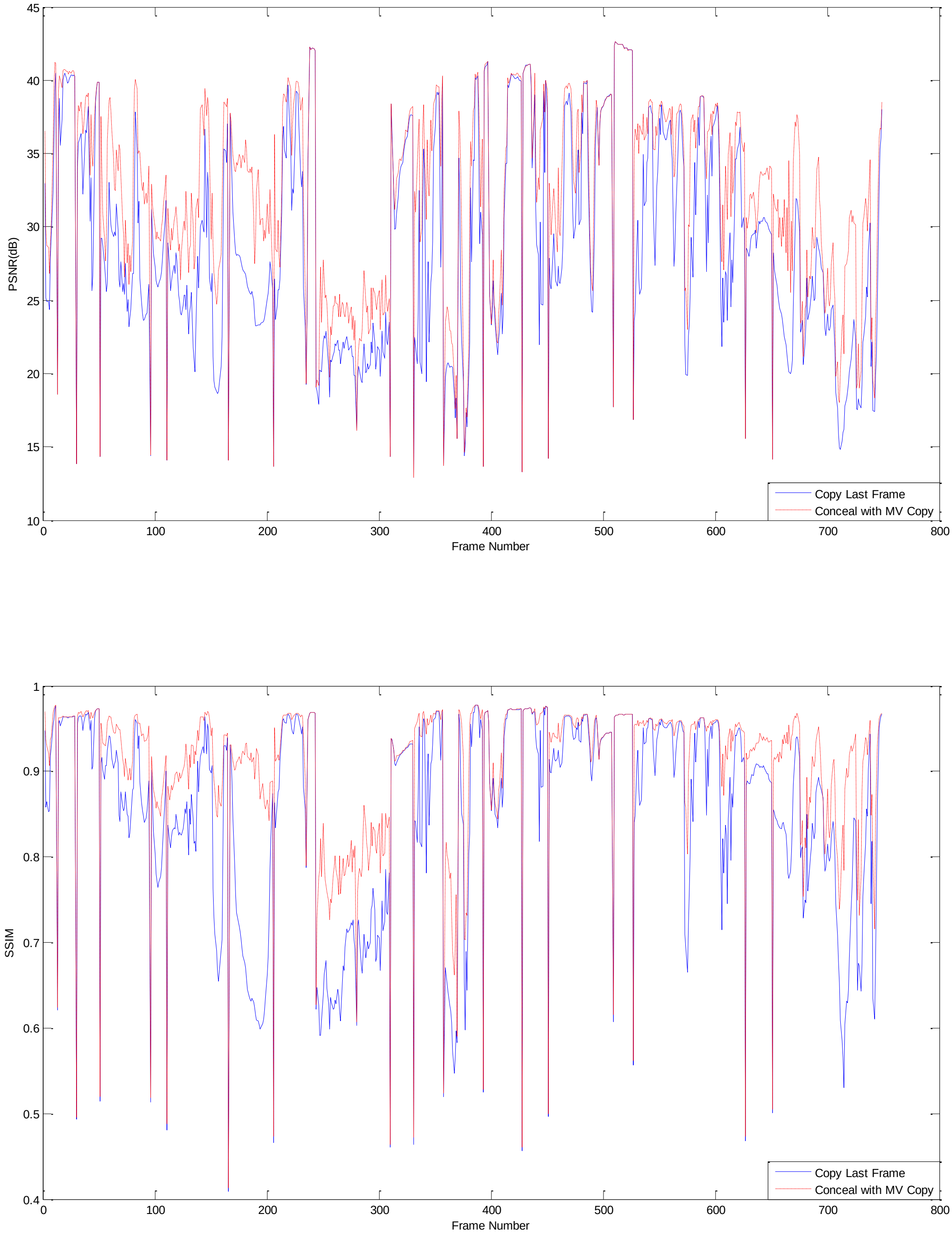
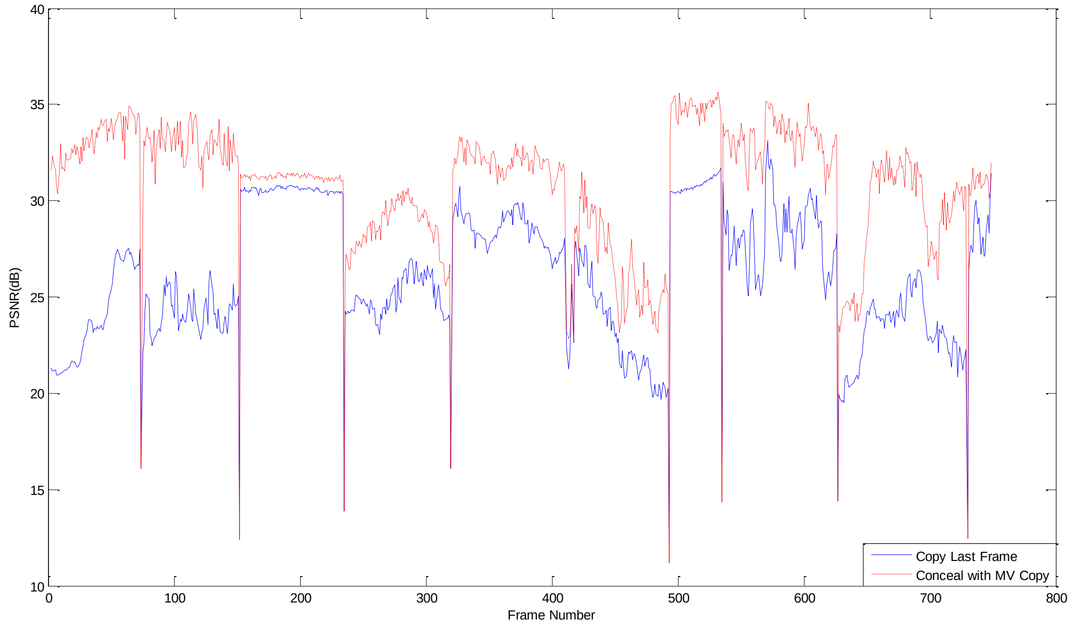


Figure 5-21: PSNR and SSIM results when left frames of “Knights Quest” video are lost and concealed with copying last left frame and copying right MVs.



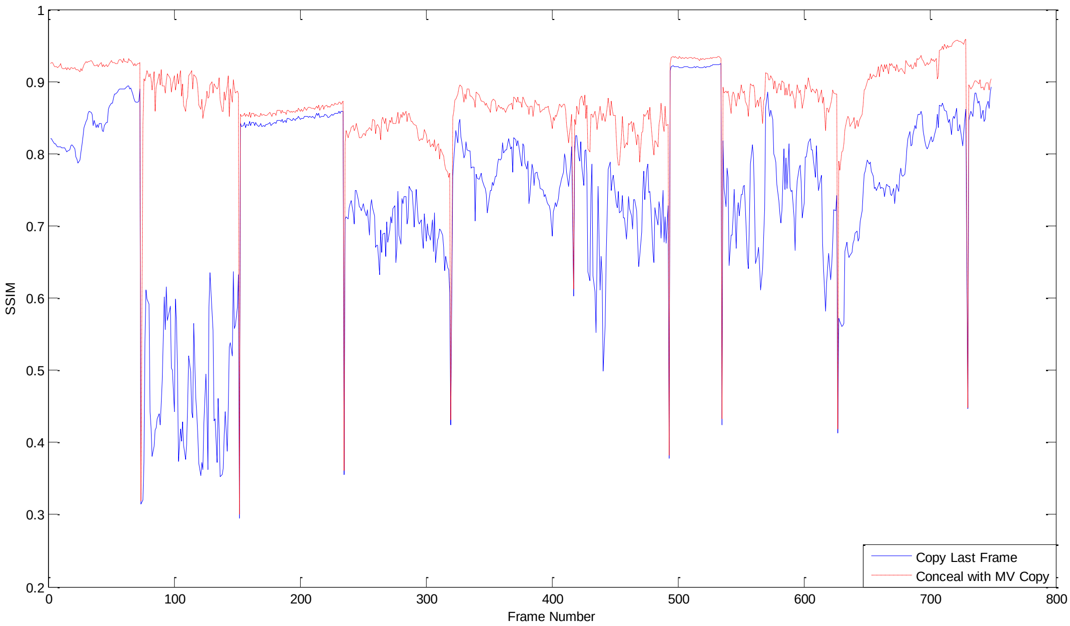


Figure 5-22: PSNR and SSIM results when left frames of "Rhine Valley Moving" video are lost and concealed with copying last left frame and copying right MVs.

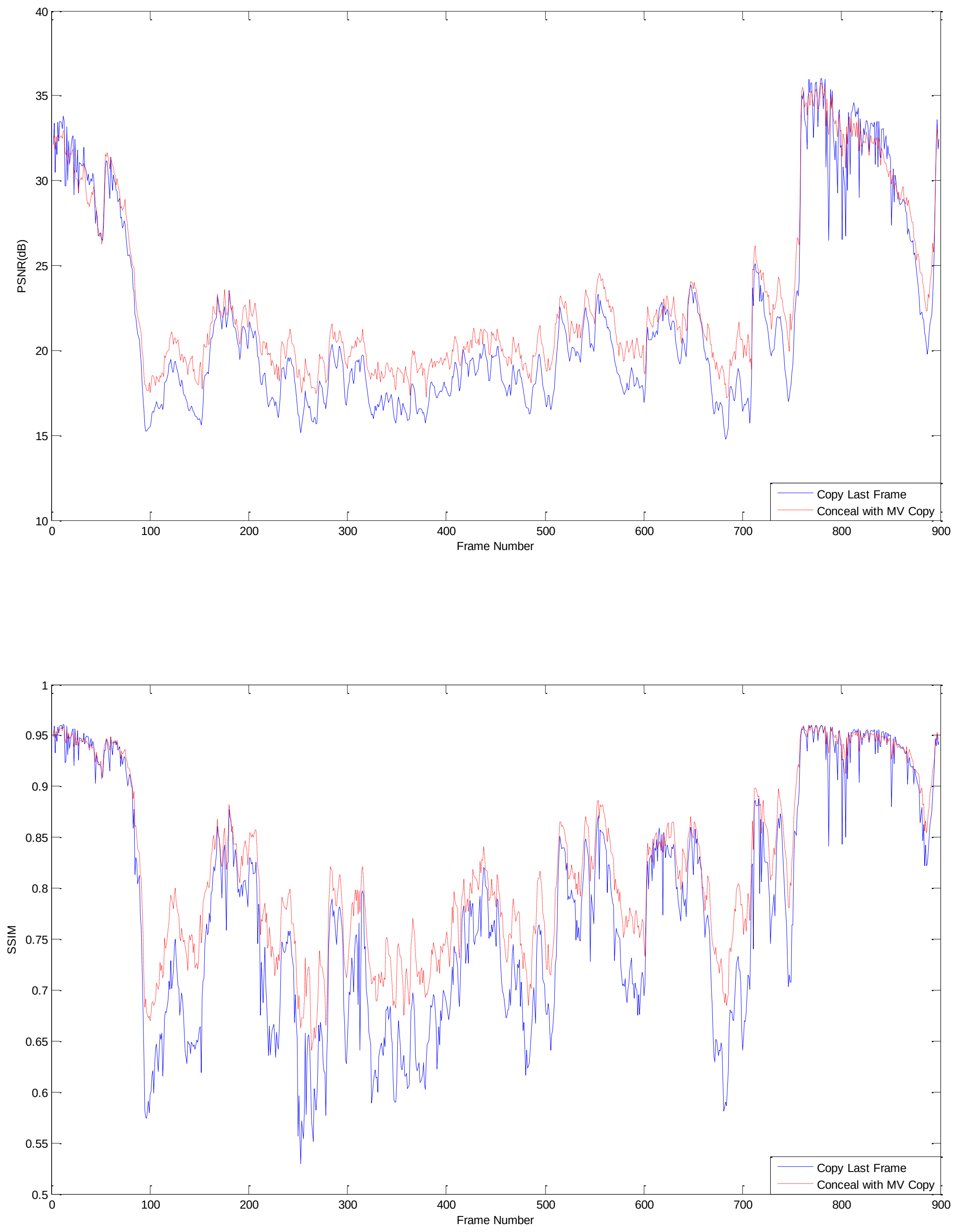


Figure 5-23: PSNR and SSIM results when left frames of "Rollerblade" video are lost and concealed with copying last left frame and copying right MVs.

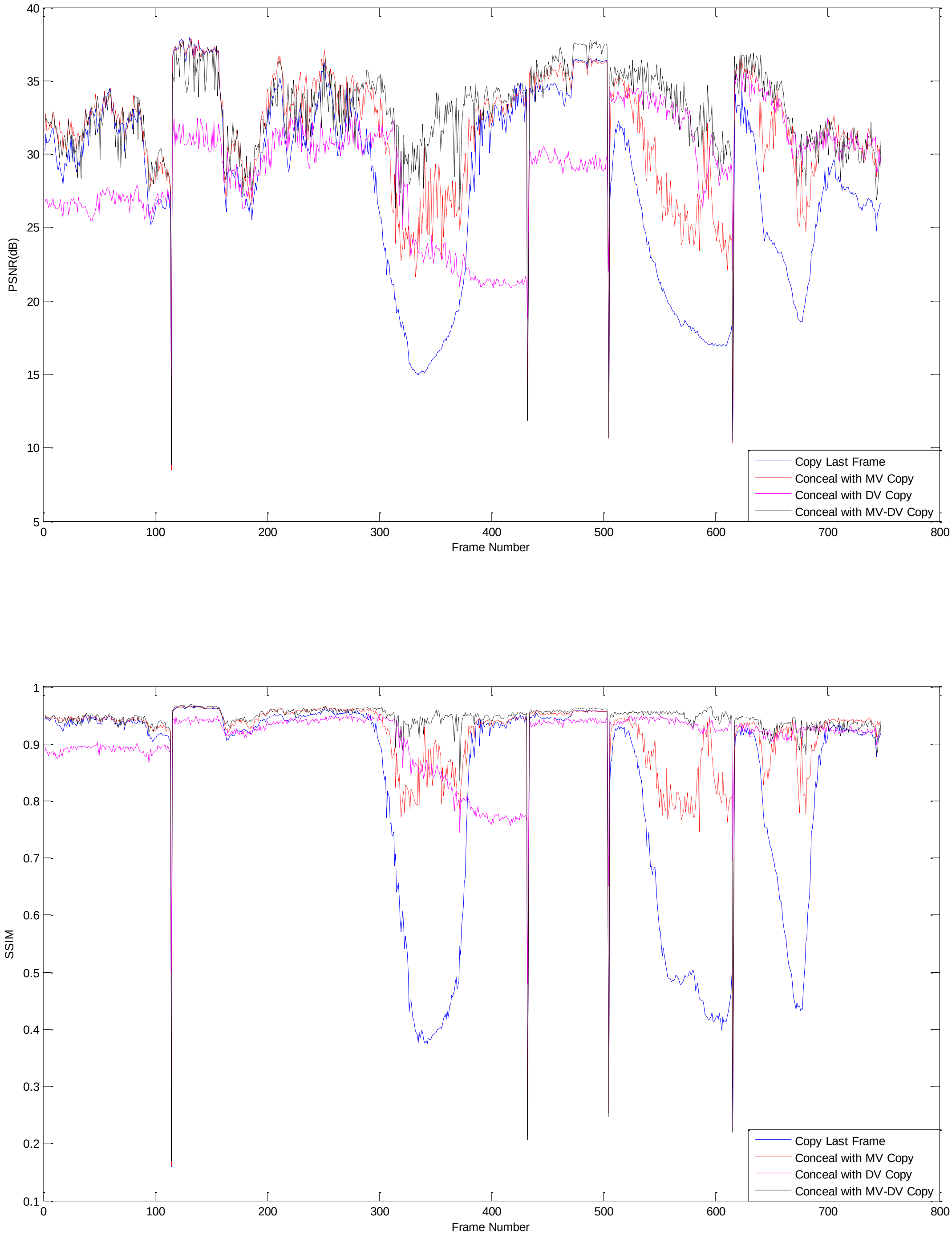


Figure 5-24: PSNR and SSIM results when right frames of "Heidelberg Alleys" video are lost and concealed with copying last right frame, copying left MVs, copying last DVs, and copying left MVs and last DVs.

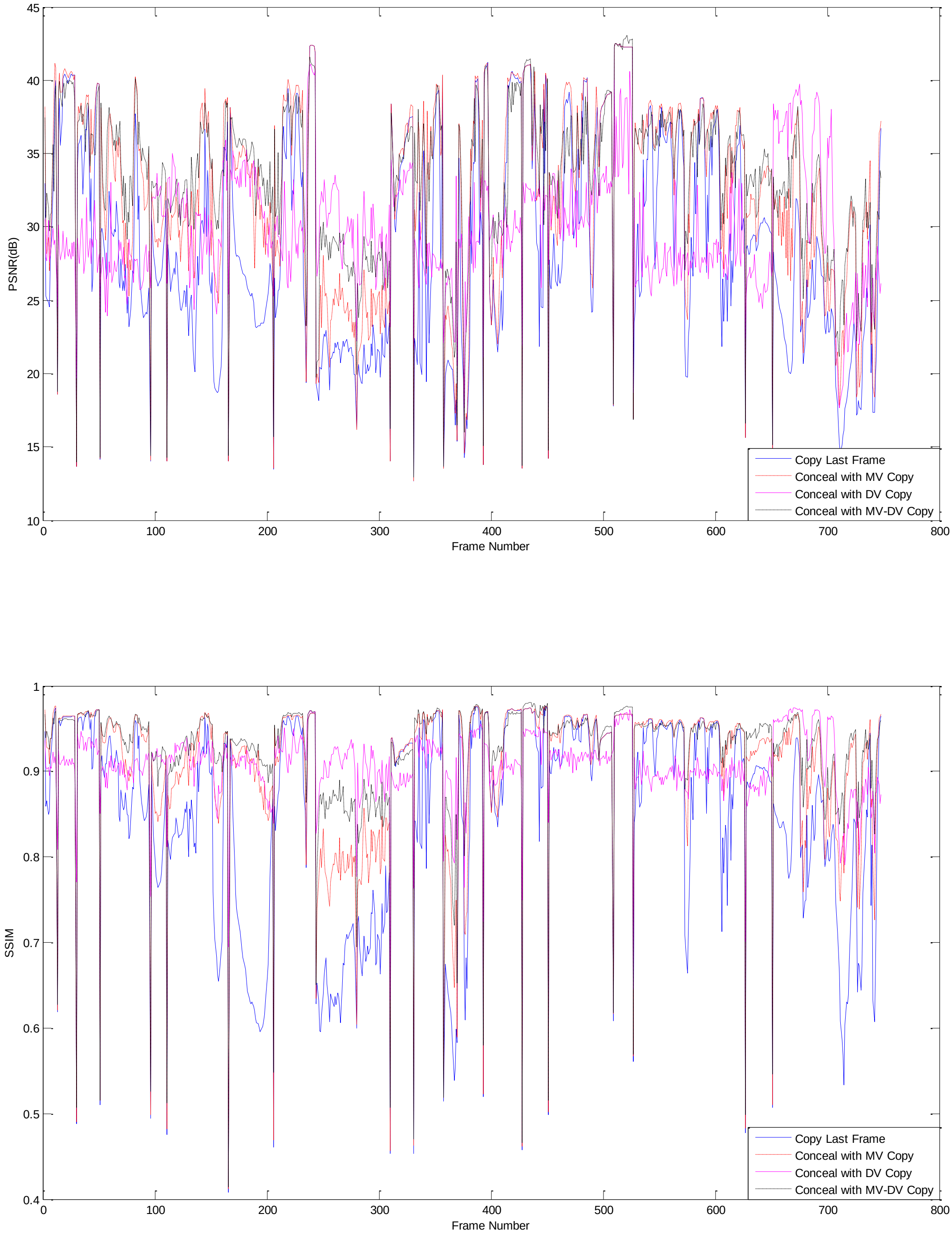


Figure 5-25: PSNR and SSIM results when right frames of "Knights Quest" video are lost and concealed with copying last right frame, copying left MVs and copying last DVs, and copying left MVs and last DVs.

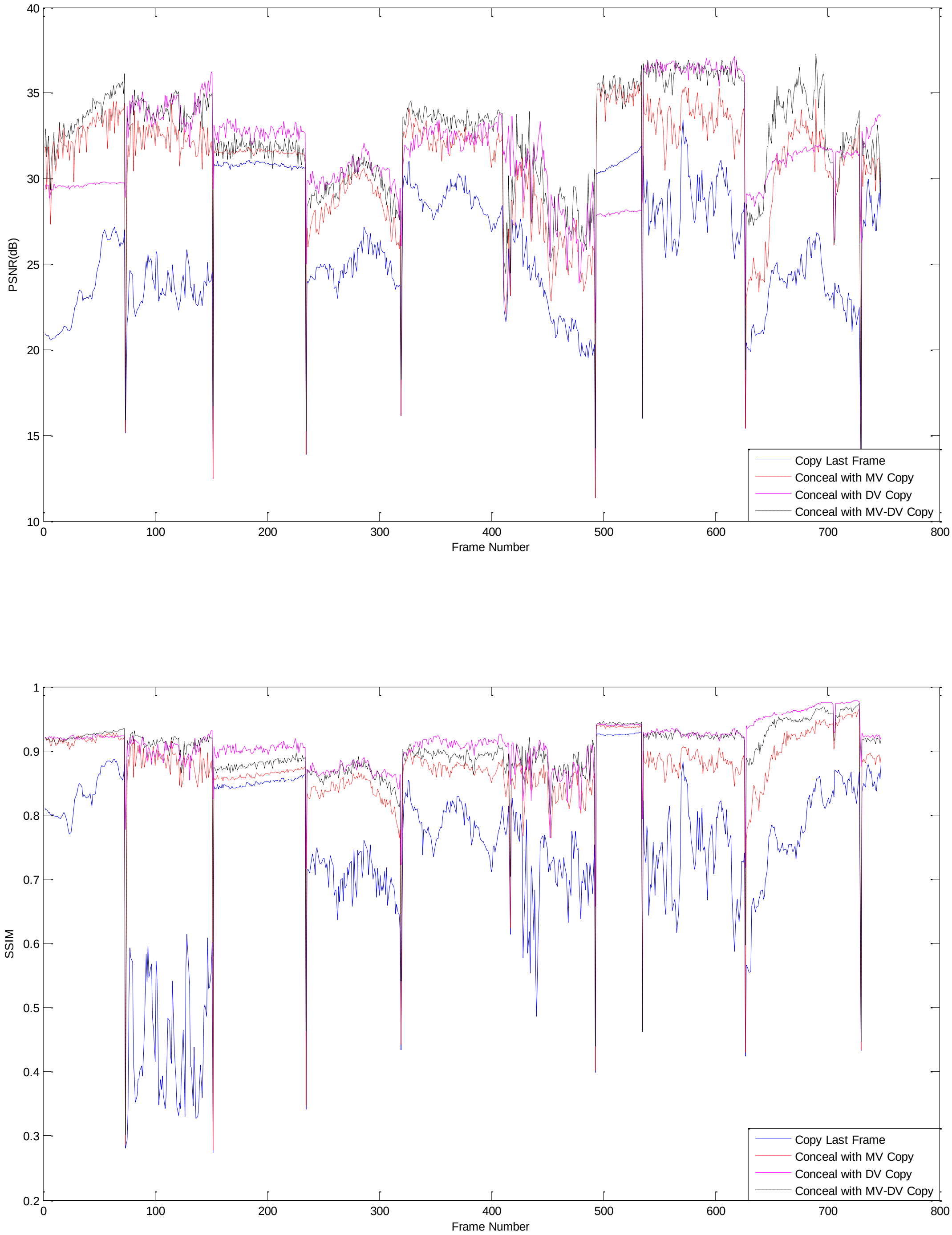
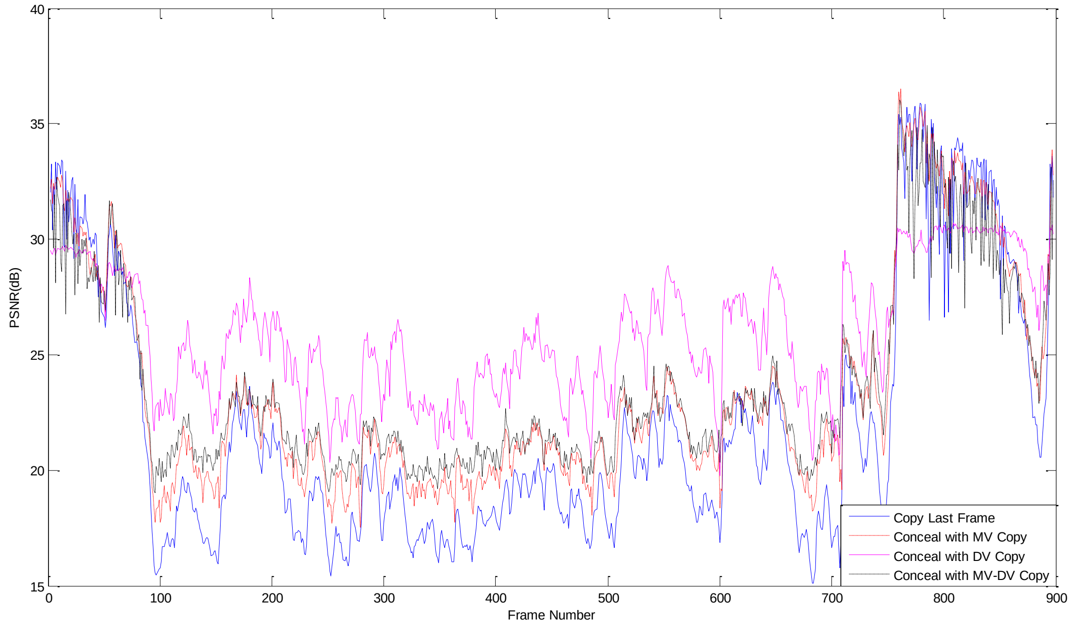


Figure 5-26: PSNR and SSIM results when right frames of "Rhine Valley Moving" video are lost and concealed with copying last right frame, copying left MVs and copying last DVs, and copying left MVs and last DVs.



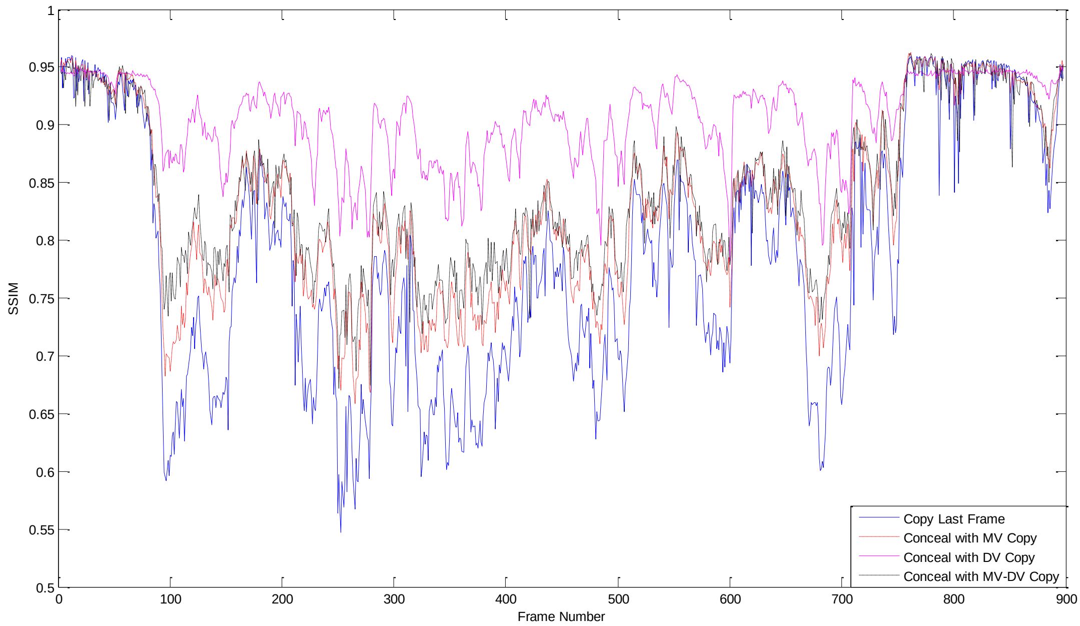


Figure 5-27: PSNR and SSIM results when right frames of "Rollerblade" video are lost and concealed with copying last right frame, copying left MVs and copying last DVs, and copying left MVs and last DVs.

The CMVC method for stereo video gives very good results when the PSNR and SSIM graphs are examined. The method gives better performance for all scenes in all of the video sequences compared to other methods. Even for the still scenes, the CMVC method gives better results than FC. For the still scenes, the FC method might be preferred because of the simplicity, however determining the type of the scene requires extra computational cost. Therefore instead of determining the scene type and choosing the concealment method, directly applying CMVC for all scenes, except scene changes, is preferable. In Table 5-4, it can be seen that, on average, the CMVC method gives 1-3 dB higher PSNR than simple FC method.

As for the performance of the DVC, the method gives better results for moving scenes. However for the still scenes, the performance of the DVC is better than the FC method for moving scenes. However, for still scenes, performance of DVC method is not better than FC. Because for still scenes, there are nonzero DVs due to the shift between views. Coding with nonzero DVs and losing residuals introduce error. However this error is smaller for the MV coded still scenes when the residuals are lost, because most of the MBs are already coded as skip MBs. The DVC method gives similar performance results for the different types of scenes. This means that for all scenes the method gives a PSNR result close to the average. This is originated from the fact that, the DVs between two views are close to each other and not much dependent to the motion in the scene. Except for the still scenes, the DVC method can be used instead of FC method. When the method is compared to CMVC method, it can be seen that the performance of DVC is not better at the still scenes. However for the moving scenes, the scene characteristics affect the results. For the camera pan, the CMVC method gives better results. It is the result of the MV uniqueness all over the scene for a camera movement. On the other hand DVC method performs better for the complex moving scenes such that many object moves in the scene independently. Since MVs at the object edges do not correspond to the correct indices, as explained before, the error increases. In most of the scenes DVC method gives good results, even better than CMVC method, but it is not appropriate to use this method without content check.

The CDVMVC method gives better results than FC for all cases. Although the DVC method does not give better performance for still scene, a combination with MVC gives better performance than FC. The combination method is superior to the CMVC method for all scenes too. As can be seen from the graphs, the CDVMVC method gives almost constant performance independent from the scene types as DVC method. Although it does not give better results than the DVC method for the scenes which contain smaller and independently moving objects, it is more reliable.

In Table 5-4, the average PSNR and SSIM results for the video sequences are given. It can be seen that CDVMVC method gives the optimum solution for all scene types.

Table 5-4: Average PSNR and SSIM results for the Frame Copy, Complementary MV Copy, DV Copy and Combined MV-DV Copy Methods

Video	Lost View	PSNR				SSIM			
		Frame Copy	MV Copy	DV Copy	MV-DV Copy	Frame Copy	MV Copy	DV Copy	MV-DV Copy
Heidelberg Alleys	Left	28,15	31,50	NA	NA	0,8213	0,915	NA	NA
Knights Quest	Left	28,89	32,30	NA	NA	0,8491	0,9060	NA	NA
Rhine Valley Moving	Left	26,06	30,82	NA	NA	0,7462	0,8736	NA	NA
Rollerblade	Left	21,74	23,07	NA	NA	0,7800	0,8210	NA	NA
Heidelberg Alleys	Right	28,11	31,56	29,02	32,91	0,8219	0,9159	0,9066	0,9430
Knights Quest	Right	28,88	32,30	29,78	33,30	0,8485	0,9057	0,9075	0,9228
Rhine Valley Moving	Right	26,11	30,94	31,76	32,45	0,7443	0,8747	0,9113	0,9009
Rollerblade	Right	21,92	23,47	25,88	23,68	0,7861	0,8315	0,9051	0,8410

In the following pictures, it can be seen that FC gives the most natural frame, where blocking effect is not present. For all other methods, there is artificial erroneous blocking effect. Also as explained before, the temporal block copy methods introduce replicas in the concealed frame. In the results of DVC and CMVDVC methods, there are horizontally misplaced blocks and block sequences, which do not exist in the results of CMVC method. In the CMVC method, the source of error is the shift of MVs between the two views, and the absence of the MVs for the intra coded blocks. For the motionless background areas, no error is introduced. However, this is not the case for the DVC method. For DVC, there is an error introduced for motionless areas too. Since there is no information for the leftmost area of the images, there is an artificial column for the DV compensated images. This is not the case for the CMVC and CDCMVC methods.

(a) (b)

(c) (d)

(e)

Figure 5-28: (a) 262nd right frame of original Rollerblade video, (b) 261st right frame of decoded Rollerblade video, (c) 262nd right frame of Rollerblade video is concealed with CMVC (d) 262nd right frame of Rollerblade video is concealed with LDVC (e) 262nd right frame of Rollerblade video is concealed with CMVDVC

CHAPTER 6

CONCLUSIONS

In this thesis, simple error concealment algorithms for full frame losses are implemented for depth plus video and stereo video representations. Although the implemented methods are tested for full frame losses, they can be applied to the partial frame losses too. It is not appropriate to use the methods for the MB error concealment, since large scale motions or zero motions are recovered well but the small object movements are not concealed successfully.

The algorithms are proposed for the low resolution systems and the systems having limited computational power. The only modification in the decoder is construction of a virtual 4x4 block synthesizer and a MV – DV buffer when a loss is discovered.

For the 2D+depth systems, it is realized that in case of a depth frame loss, use of FCBR method gives better performance, whereas in case of a left frame loss, it does not yield very good results. When the computational complexity of a renderer is considered, it is more logical to directly copy the last right frame. The CMVC for 2D+depth systems does not give superior results for both left frame and depth frame loss cases. The depth frames are not textured, so the temporal prediction may not give the precise flow of motion. Also the MBs of the depth frames are coded mostly with either intra frames or residuals. Therefore usage of the depth MVs does not give good results. Use of the left MVs for the depth frame concealment does not give good performance for the scenes where the objects moves towards or away from the camera, because the depth sequence introduces illumination changes, where left MVs cannot recover.

For the stereoscopic systems, it is deduced that using CMVC method always gives better results than FC method. For left frame concealment, using CMVC method for all scenes instead of FC method is preferable. For the MVC coding structure, since the right frame is decoded using left frame and last right frames, only limited number of MVs can be used for the concealment of left frame. Therefore the CMVC method might not be appropriate for the left frame concealment in MVC structure. Using monoscopic concealment schemes might be more logical for the left frame loss cases.

For the right frame loss cases, there are three possible concealment scheme for MVC structure. The CMVC method performs well for all scene types. The performance of the CMVC is absolutely better than performance of FC when the scene is moving, however they are similar for the still scenes. Using the DVC method is preferable for the moving scenes however it is not consistent to use this method for still scenes. For the moving scenes, DVC method gives worse results than MVC for camera pan and large object movements. In the MVC structure, since both MV and DV compensated blocks exist in the previous frame, the number of the DVs might be too few. Since histogramming introduces errors, a combination of the MV and DV usage is introduced. The CMVDVC method gives better results than FC for all scenes. Also the average performance is better than CMVC method. Although it is not better than DVC method for the scenes containing small moving objects, since the performance is more deterministic for all scene types, it can be preferred.

PSNR and SSIM results are improved compared to the FC method for the implemented methods for at least special scene types. However the visual quality is not much improved. Simple FC method always gives natural scenes, whereas the proposed methods cause artifacts in the scenes. Mostly blocking effects occur in the scenes, and replicas of the moving objects come out, because not for all blocks accurate MVs or DVs are saved.

In DVC method, the intra coded blocks are compensated with an estimated DV. This method introduces wrong DV compensations, and extra artifacts and erroneous translations in the resulting frame.

In hierarchical coding structure, frames can be compensated with the previously decoded frames within the same GOP. However, in this thesis hierarchical coding structure is not fully employed, where the B frames and multiple references are not utilized. For future work, the algorithms can be implemented with full hierarchical coding, where the MV and DV buffer need to preserve the reference number to compensate with the correct reference frame. Also for multiple reference frames, the MV-DV Buffer counts need to be increased.

To conceal a lost frame, a low cost solution is implemented in this thesis. Whereas increasing computational complexity, the MV and DV buffers can be filtered or interpolated. In addition, the current implementation does not consider the scene characteristics. The method can be modified such that, the motion vectors of correctly received complementary view can be used to define the scene characteristics i.e. fast motion-slow motion, or camera panning. Than the selection between DV copy, MV copy and frame copy can be made adaptively to obtain the best results.

REFERENCES

[1] "ITU-T Rec. H.264-ISO/IEC 14496-10 AVC, Advanced Video Coding for Generic Audiovisual Services," March 2009.

[2] ITU-T Recommendation H.264, "Advanced video coding for generic audiovisual services", November 2007.

[3] A. Vetro, P. Pandit, H. Kimata, A. Smolic, and Y.-K. Wang, Joint draft 8 of multiview video coding, Hannover, Germany, Joint Video Team (JVT) Doc. JVT-AB204, Jul. 2008

[4] ITU-T and ISO/IEC JTC 1, Advanced video coding for generic audiovisual services, ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), 2010.

[5] M. Wada, "Selective recovery of video packet loss using error concealment," IEEE J. Select. Areas Commun. vol. 7, pp. 807–814, June 1989.

[6] R. Aravind, R. Civanlar, and A. R. Reibman, "Packet loss resilience of MPEG-2 scalable video coding algorithms," IEEE Trans. Circuits Syst. Video Technol., vol. 6, pp. 426–435, Oct. 1996.

[7] E. Ayanoglu, P. Pancha, A. R. Reibman, and S. Talwar, "Forward error control for MPEG-2 video transport in a wireless ATM LAN," in Proc. ICIP '96, Lausanne, Switzerland, Sept. 1996, pp. II 833–836.

[8] 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.

[9] L. H. Kieu and K. N. Ngan, “Cell-loss concealment techniques for layered video codecs in an ATM network,” IEEE Trans. Image Processing, vol. 3, pp. 666–677, Sept. 1994.

[10] Y. Wang, Q.-F. Zhu, and L. Shaw, “Maximally smooth image recovery in transform coding,” IEEE Trans. Commun., vol. 41, pp. 1544–1551, Oct. 1993.

[11] W. Zhu and Y. Wang, “A comparison of smoothness measures for error concealment in transform coding,” in Proc. SPIE Conf. Visual Communication and Image Processing, Taipei, Taiwan, 1995, vol. II, pp. 1205–1214.

[12] W. Kwok and H. Sun, “Multi-directional interpolation for spatial error concealment,” IEEE Trans. Consumer Electron., vol. 39, pp. 455–460, Aug. 1993.

[13] H. Sun and W. Kwok, “Concealment of damaged block transform coded images using projections onto convex sets,” IEEE Trans. Image Processing, vol. 4, pp. 470–477, Apr. 1995.

[14] D. L. Robie, and R. M. Mersereau, "The use of Hough transforms in spatial error concealment," in Proc. of IEEE ICASSP 2000, Istanbul, Turkey, pp. 2131-2134, 2000

[15] S. Aign, "Error Concealment Improvements for MPEG-2 Using Enhanced Error Detection and Early Re-Synchronization," IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, 1997

[16] S. S. Hemami and T. H.-Y. Meng, “Transform coded image reconstruction exploiting interblock correlation” IEEE Trans. Image Processing, vol. 4, pp. 1023–1027, July 1995.

[17] S. Aign and K. Fazel, “Temporal & spatial error concealment techniques for hierarchical MPEG-2 video codec,” in Proc. Globecom’95, pp. 1778–1783.

[18] S. Tsekeridou, F. A. Cheikh, M. Gabbouj, and I. Pitas, "Motion field estimation by vector rational interpolation for error concealment purposes", IEEE Int. Conf. on Acoustics, Speech and Signal Proc. 1999.

[19] L. Atzori, F. Natale and C. Perra, "A Spatio-Temporal Concealment Technique Using Boundary Matching Algorithm and Mesh-Based Warping (BMA–MBW)" IEEE Trans. Multimedia, vol. 3, no. 3, pp. 326-337, Sep. 2001.

[20] Y.-C. Lee, Y. Altunbasak and R. M. Mersereau, "Multiframe error concealment for MPEG-coded video delivery over error-prone networks", IEEE Trans. Image Processing, vol. 11 no. 11, pp. 1314–1331, Nov. 2002.

[21] M-J. Chen, C-S. Chen, M-C. Chi, "Temporal error concealment algorithm by recursive block-matching principle", IEEE Trans. Circuits and Systems for Video Technology, vol. 15, no. 11, pp. 1385-1393, Nov 2005.

[22] S. Belfiore, M. Grangetto, E. Magli, and G. Olmo, "Concealment of whole-frame losses for wireless low bit-rate video based on multiframe optical flowestimation," IEEE Trans. Multimedia, vol. 7, no. 2, pp. 316–329, Apr. 2005.

[23] Q. Peng, T. Yang, and C. Zhu, "Block-based temporal error concealment for video packet using motion vector extrapolation," in Proc. IEEE Int. Conf. Communications, Circuits and Systems and West Sino Expositions, Jun. 2002, vol. 1, pp. 10–14.

[24] Y. Chen, K. Yu, J. Li, and S. Li, "An error concealment algorithm for entire frame loss in video transmission," Proc. IEEE Picture Coding Symp., 2004.

[25] J. Zhou, B. Yan and H. Gharavi, "Efficient motion vector interpolation for error concealment of H.264/AVC", IEEE Trans. Broadcasting, vol. 57, no. 1, pp. 75-80, Mar. 2011.

[26] S. Knorr, C. Clemens, M. Kunter and T. Sikora, "Robust concealment for erroneous block bursts in stereoscopic images", Proc. of the 2nd International Symp. on 3DPVT, 2004.

[27] L. Pang, M. Yu, W. Yi, G. Jiang, W. Liu and Z. Jiang, "Relativity analysis-based error concealment algorithm for entire frame loss of stereo video", IEEE ICSP, 2006.

[28] T.-Y. Chung, S. Sull, and C.-S. Kim, “Frame loss concealment for stereoscopic video based on inter-view similarity of motion and intensity difference”, IEEE 17th International Conference on Image Processing, Sept. 2010

[29] J. Carreira, L. Pinto, N. Rodrigues, S. Faria, P. Assuncao, “Subjective assessment of frame loss concealment methods in 3D video”, 28th Picture Coding Symposium, Dec. 2010

[30] B. Yan, “A novel H.264 based motion vector recovery method for 3D video transmission” IEEE Trans. Consumer Electronics, vol. 53, no. 4, pp. 1546-1552, Nov. 2007.

[31] Y. Liu, J. Wang and H. Zhang, “Depth image-based temporal error concealment for 3-D video transmission”, IEEE Trans. Circuits and Systems for Video Technology, vol. 20, no. 4, pp. 600-604, Apr. 2010

[32] A. Ali, H. A. Karim, N. A. M. Arif, A. Sali, “Depth image-based spatial error concealment for 3-D video transmission”, Proc. 2010 IEEE Student Conference on Research and Development, 2010

[33] C. T. E. R. Hewage, S. T. Worrall, S. Dogan, A. M. Kondoz, “A novel frame concealment method for depth maps using correponding colour motion vectors”, 3DTV-CON 2008

[34] C.T.E.R. Hewage, S. Worrall, S. Dogan, and A.M. Kondoz, “Frame concealment algorithm for stereoscopic video using motion vector sharing”, IEEE International Conference on Multimedia and Expo, Hannover, Germany, Jun. 2008

[35] http://www.dongleware.de/, Last Accessed on 10/09/2009

[36] http://www.redstarstudio.co.uk/, Last Accessed on 10/09/2009

[37] http://www.cinovent.de/, Last Accessed on 10/09/2009

[38] http://www.hhi.fraunhofer.de/, Last Accessed on 10/09/2009

[39] Feldmann, I. Atzapadin, N, Schreer, O. Pujol-Acolado, J.-C. Landabaso, J.-L. Escoda, O. D. “Multi-View Depth Estimation Based on Visual-Hull Enhanced Hybrid Recursive Matching for 3D Video Conference Systems” ICIP 2009, Kairo, Egypt, November 2009

[40] Merkle, P. Wang, Y. Müller, K. Smolic, A. and Wiegand, T. “Video plus depth compression for mobile 3D services," 3DTV Conference, Potsdam, Germany, 2009

[41] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity", IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, Apr. 2004.

[42] https://ece.uwaterloo.ca/~z70wang/research/ssim/, Last Accessed on 29/07/2011