國立交通大學

電子工程學系　電子研究所碩士班

碩 士 論 文

一個可應用在 MPEG-21 多媒體測試平台上之具有時間/空間適應性錯誤補償的 MPEG-4 解碼器

An MPEG-4 Simple Profile Decoder with Adaptive Spatial/Temporal Error Concealment for MPEG-21 Multimedia Test Bed

研 究 生：黃名彥

指導教授：　蔣迪豪 博士

中 華 民 國 九 十 三 年 六 月

一個可應用在 MPEG-21 多媒體測試平台上之時間/空間適應性錯誤補償的 MPEG-4 解碼器

An MPEG-4 Simple Profile Decoder with Adaptive Spatial/Temporal Error Concealment for MPEG-21 Multimedia Test Bed

研究生: 黃名彥　　　　Student: Ming-Yen Huang

指導教授: 蔣迪豪　　　　Advisor: Tihao Chiang

國 立 交 通 大 學

電子工程學系 電子研究所碩士班

碩 士 論 文

A Thesis

Submitted to Department of Electronics Engineering & Institute of Electronics

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Electronics Engineering

June 2004

HsinChu, Taiwan, Republic of China

中華民國九十三年六月

一個可應用在MPEG-21多媒體測試平台上之時間/空間適應性 錯誤補償的MPEG-4解碼器

研究生: 黃名彥 指導教授: 蔣迪豪 博士

國立交通大學

電子工程學系 電子研究所碩士班

摘要

為了在 MPEG-21 多媒體通通存取(Universal Multimedia Access)架構中實現一個強韌的視訊內容傳送機制，所有的資訊包含結構、動作訊息、網路狀況、和接收端能力等等都會被考量來提供使用者最好的視訊品質。因此，視訊內容傳送技術必須考量結構、動作訊息、和網路狀況等因素。具有不同錯誤回覆方法的接收端將呈現了不同的視訊內容品質。本論文著墨於在解碼端強韌的容錯力與錯誤回復力。對於在網際網路或無線網路上傳輸的視訊串流，我們提出了一個容錯的解碼器來抵抗網路封包遺失，並且回復解碼程序。除此之外，我們提出一個適應性的時間/空間錯誤補償機制來改善畫面的品質。在錯誤補償之後，一個簡單的低頻濾波器被用來減少因補償之後造成的塊狀效果。在模擬結果中，我們分析每一個補償工具。模擬的結果顯示出我們能夠有效地改善視訊串流在網際網路或無線網路上的視覺品質效果。比較適應性的時間/空間補償方法與在編碼端使用重新同步記號和在解碼端使用零向量及空間複製的補償方法,我們提出的方法可以得到3.65至9.71dB的最大訊躁比(Peak Signal-to-Noise Ratio)之增益。最後，我們整合這個具有容錯能力與錯誤補償方法的解碼器到 MPEG-21 多媒體測試平台來測試這個錯誤補償方法在實際網路狀況的效果。

An MPEG-4 Simple Profile Decoder with Adaptive Spatial/Temporal Error Concealment for MPEG-21 Multimedia Test Bed

Student: Ming-Yen Huang Advisor: Dr. Tihao Chiang

Department of Electronic Engineering &
Institute of Electronics
National Chiao Tunrg Univesity

Abstract

To realize a robust video content delivery scheme under MPEG-21 UMA (Universal Multimedia Access), all information covering texture, motion information, network conditions and terminal capabilities are used to serve consumers with the best visual quality. In our technique, the texture, motion information and network conditions are considered. The terminals with various error recovery methods have shown different quality of video contents. This thesis will focus on the error robustness and error recovery capabilities of the media decoder. For streaming video over the Internet and wireless channels, we propose a resilient decoder against packet loss to resume the decoding process. In addition, an adaptive temporal/spatial concealment scheme is proposed to further improve the picture quality of reconstructed video. After concealment, a simple smooth filter is used to reduce blocking artifact. In simulation, the impact of each error concealment tool used is analyzed. Our simulation results show that we can significantly improve subjective quality for video streaming over the Internet and wireless channels. The adaptive temporal/spatial concealment can get 3.65-9.71 dB gain in PSNR as compared to that of the streaming scheme that employs resynchronization marker at the encoder and uses zero motion vector and spatial copy concealment at the decoder. Finally, the error resilient and error concealment decoder is integrated into MPEG-21 part-12 multimedia test bed to evaluate the on-line performance of adaptive concealment scheme under the real network conditions.

誌謝

在研究所的這兩年當中，首先要感謝指導教授蔣迪豪教授，由教授在帶實驗室的衝勁、對學生的要求與期許、以及協助就業的過程當中，從教授身上學到、感受到了很多的東西，在此對蔣教授獻上最大的感謝之意。

感謝所有其他實驗室的學長與同學，尤其是王俊能與王士豪學長，嚴謹認真的研究態度，研究上需要注意的事項，與其他很多的協助與引導，都給了我在研究所生活當中很好的磨練。而同在五樓實驗室的的孝強、耀諄、子良、俊安、振瑋、岳賢、子瀚也在學習過程中提供了很多的幫助與一起學習的回憶。其他通訊電子暨訊號處理實驗室的學長姐、同學、學弟妹，大家營造出了一個氣氛融洽的研究環境，陪伴我走過這兩年的研究生活。

感謝從小辛苦把我扶養長大的父母，有你們的用心，你們的付出，才能夠造就我今日的成就。我也會再接再勵，來報答你們的恩情。

其他還要感謝很多的親朋好友，有你們的扶持與鼓勵，讓我渡過了很多的艱難與考驗。最後，僅以這篇論文，獻給所有陪伴我走過這段日子的人，謝謝。

黃名彥

民國九十三年六月 於新竹

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Recently, multimedia communications systems are more and more popular. Through transmission network, multimedia systems can serve users plenty of media contents, which have become a business or product that can make life more colorful, fast, and convenient. Multimedia technology provides the different players in the multimedia value and delivery chain with an excess of information and services. Access to information and services can be provided with ubiquitous terminals and networks from almost anywhere at anytime. However, no solution enables all the heterogeneous communities to interact and interoperate with one another. To address the interaction and interoperation, MPEG-21 Universal Multimedia Access (UMA) defines a description of how various elements including networks, terminals, user preference, and natural environments, etc. can fit together. MPEG-21 UMA defines a multimedia framework to facilitate transparent and augmented use of multimedia resources across a wide range of networks and devices employed by different communities. In additional, UMA allows content adaptation according to terminal capabilities. Each terminal has various abilities including media format, error recovery, and so on. For content adaptation according to terminal capabilities and network conditions, binary resource adaptation is specified by MPEG-21 part-7 Digital Item Adaptation.

For video content delivery, video coding standards are demanded. For compact storage and efficient transmission of video contents, video coding standards including MPEG-1/2/4 and H.263/H.264 have been widely used. Based on motion information, temporal or spatial correlation, video coding standards can compress the video data to much smaller size for the limited storage space or transmission bandwidth. Without compression, we may spend a huge amount of transmission bandwidth to deliver video. For example, for a sequence of duration 10 minutes, 30 frames per second, CIF resolution (352 * 288) and one byte for each component of RGB color model, we need to transmit 30*60*10*352*288*3 bytes. With MPEG-4 video coding at 800 kilo-bits/sec that can provide acceptable visual quality, only 30*60*800/8 bytes are needed to be transmitted via the channels, which has reduction of

30412 times.

Fig. 1. Video streaming flow over the network.

For video application, video storage, VCD or DVD may have some man-made damages to cause bits error. For more and more applications that access and retrieve video content over the Internet or wireless channels, the transmission errors by the fading channel and time-variant network conditions cover packet loss, delay and jitter. For example, Fig. 1 shows a simplified video streaming framework. In delivering the bitstreams over the transmission channels, various types of distortion will discard parts of the transmitted bitstream, which degrades the visual quality of receiving video content for consumers at the receiving end. Thus, a error resilient and robust streaming system is strongly demanded for providing a satisfactory quality of service for network users at the occurrence of any transmission errors. To evaluate the quality of service for an error resilient decoding system, MPEG-21 part-12 Multimedia Test Bed for Media Delivery is used in this thesis.

Error resilience techniques have been developed and applied to various multimedia players. The existing error resilience techniques can be categorized into 4 major stages including error detection, error recovery, error localization and error concealment as shown in Fig. 2. The error detection process detects the syntactic and semantic errors to prevent the failure of the decoder system by any exception. The error detection process is critical for the subsequent three steps. The error recovery process takes over the detected errors to recover the correct process and continue decoding. The error localization locates the error position and a proper error concealment approach is used to amend the visual quality. Thus, with the error detection, the error recovery, the error localization and the error concealment, the resilient video decoding system can be used for the multimedia applications under the error prone environment. To demonstrate an error resilient decoding, we realize a crash proof and error concealment decoder based on Simple Profile of MPEG-4 part-2 visual video coding. Our error resilience approaches that are complaint with block-based video coding standards are applicable to other block-based coding schemes with some adaptation.

For video streaming over the Internet and wireless channels, the bitstreams may be corrupted by packet loss in the channels. To recover the video contents from the degradation by packet loss during transmission, the error resilient decoding process and the robust bitstreams are required. The error recovery methods utilize all the useful information available at the receiver to resynchronize decoding processes. For improving the received video contents to a proper visual quality, we present an adaptive temporal/spatial error concealment scheme for delivering video contents over the Internet and wireless channels. In addition to error recovery from the packet loss, our resilient approaches can conquer random bit error. In order to simulate our method on the wireless transmission and different kinds of real-time streaming network conditions, we use MPEG-21 part-12 Multimedia Test Bed for Media Delivery to simulate the on-line performance under real network conditions.

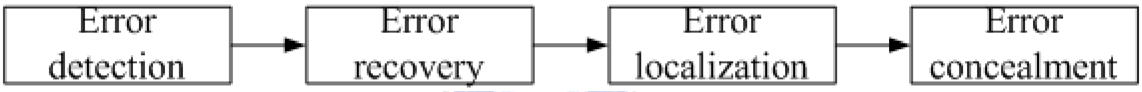


Fig. 2. Error resilience decoding process.

Error concealment methods can basically be classified into spatial error concealment and temporal error concealment. The simple application is to use the spatial error concealment and temporal error concealment to recover the corrupted I-VOP and P-VOP respectively. To further improve the visual quality, some methods have used both concealment approaches to recover all VOPs. The innovation of our proposed error concealment method is that we use some evaluations on each macroblock (MB) or each block to choose better concealment from spatial and temporal error concealment. Block-based error concealment can fit video properties more than frame-based error concealment [24]. With concealed frames, in order to remove the visible blocking effect, a smoothing filter is adopted. The simple flow of proposed error concealment process is as shown in Fig. 4. Compared with the simplest concealment method which uses resynchronization marker at encoding end and zero motion / spatial copy at decoding end, we can get visual quality improvement by 2.66dB-11.93dB in peak signal-to-noise ratio (PSNR).

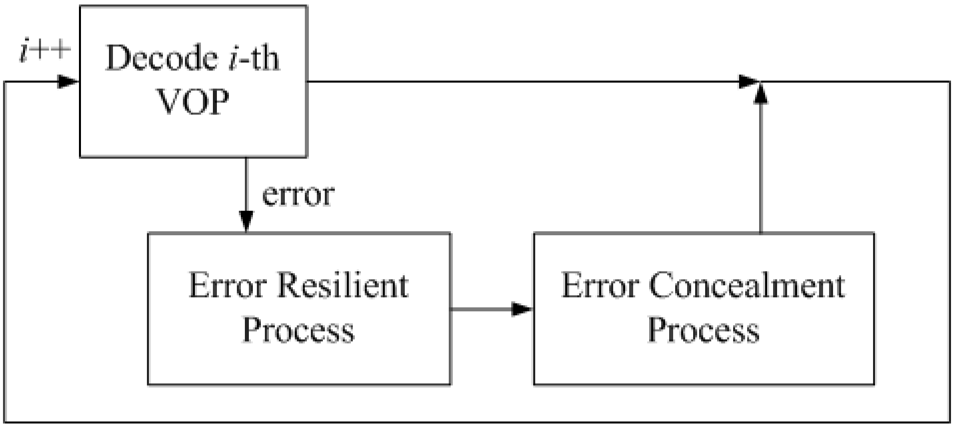


Fig. 3. Error resilient decoding system

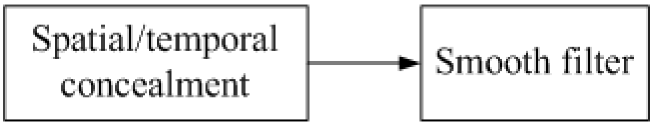


Fig. 4. Proposed error concealment scheme.

The organization is described as follows. Chapter 2 introduces MPEG-4 video coding systems and summarizes error resilience tools of MPEG-4 Simple Profile video standard and existing error concealment methods. In additional, MPEG-21 part-7 Digital Item Adaptation and part-12 Multimedia Test Bed for Media Delivery are introduced for advanced evaluation of our error recovery techniques. Chapter 3 presents the error resilient decoding system with concealment algorithm and smooth filter for providing better quality of service when any packet loss occurrence during the transmission. For real-time decoding process, we have tried to speed up the decoding rate with hardware acceleration. Chapter 4 gives the experimental results using the proposed error concealment algorithms for both off-line testing and on-line simulations on MPEG-21 part-12 multimedia test bed. Chapter 5 draws the conclusions.

Chapter 2

Error Resilience Techniques for MPEG-4 Video Coding Standard

In this chapter, we describe related works on constructing a streaming video system and error resilience techniques. For realizing MPEG-4 video coding systems, error resilience tools of MPEG-4 Simple Profile video standard and existing error recovery methods will be introduced. In additional, MPEG-21 part-7 Digital Item Adaptation and part-12 Multimedia Test Bed for Media Delivery are introduced for performance evaluation of error recovery techniques.

2.1 Overview of MPEG-4 Video Coding Standard

MPEG-4 video coding standard provides a rich set of tools for the coding of audio-visual objects. We will provide an overview of natural visual coding tools in MPEG-4 video specification. In addition, error resilience tools defined in MPEG-4 video specification consisting of resynchronization markers, data partition, and reversible variable length coding (RVLC) are introduced.

2.1.1 Video Communications using MPEG-4

MPEG-4 video coding standard is developed to provide users a new level of performance for various video communications applications covering Internet streaming and mobile multimedia applications. For these applications, both device power and channel bandwidth are very limited. For handheld devices using the batteries, the coding process of video contents is sped up for minimizing power consumption. Thus, it’s still an important issue to realize a low-complexity real-time software or hardware video encoder and decoder. To address issues of low complexity and minimal power consumption, many fast algorithms are proposed to reduce the encoding and decoding complexity.

An MPEG-4 encoder as shown in Fig. 5 consists of six common modules in encoding process. The six encoding modules are the discrete cosine transform (DCT), motion estimation, motion compensation (MC), zigzag scan, quantization (Q), and variable length coding (VLC). With the DCT, a spatial domain data with high inter-pixel correlation is mapped into spectral coefficients that are approximately uncorrelated. The nearly uncorrelated spectral coefficients are then coded with the zigzag scan, Q and VLC to produce the coded bitstream. The coded bitstream is synthesized into a reconstruction of the spatial domain input data via the decoding modules introduced in Fig. 6, which combines the inverse operation of the DCT, zigzag scan, Q and VLC. The reconstructed data are stored and further used for removing temporal redundancy between the synthesized data and the immediately next input video signal. With the reconstructed data, temporal redundancy can be removed by finding the best approximation under a specified distortion measure for each block of the current input data.

An MPEG-4 decoder as shown in Fig. 6 consists of five modules including the variable length decoder (VLD), inverse zigzag scan, inverse quantization (Q^{-1}), inverse discrete cosine transform (IDCT) and motion compensation (MC). With the decoding modules and the bitstreams, the video contents are delivered to users.

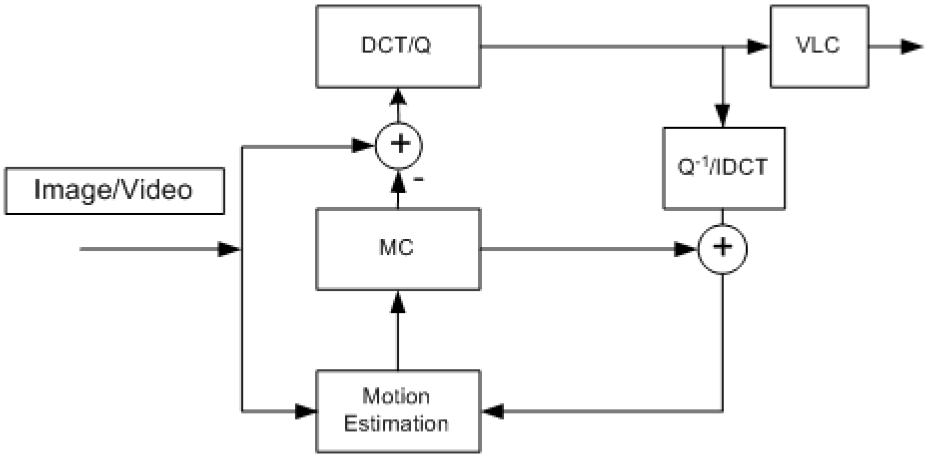


Fig. 5. Coding modules of an MPEG-4 Simple Profile encoder.

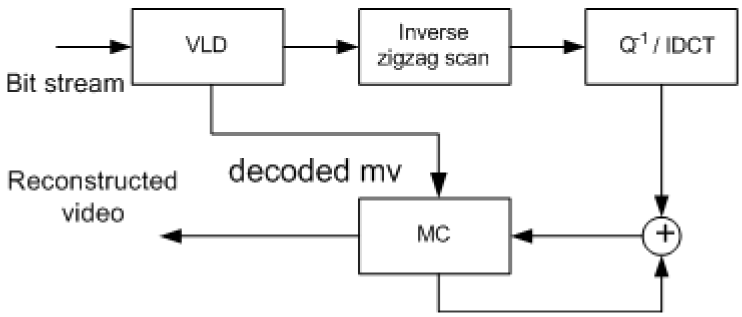


Fig. 6. Coding modules of an MPEG-4 Simple Profile decoder.

2.1.2 Overview of MPEG-4 Error Resilience

MPEG-4 compression standard is to facilitate video streaming or multimedia communications over the Internet and wireless channels. During the video data transmission, there could be error occurrence due to channel fading. For a guarantee quality of service, error resilience tools are included in the standard and take the important place to serve better video contents to end users.

2.1.2.1 Error Type and Error Resilience Process

The transmission errors are typically caused by either random bit error or packet loss. The burst bit error and packet loss without resynchronization markers can be categorized as the special case of the random bit error. The random bit error could be checked with syntactic and semantic errors. In cast of the packet loss with resynchronization markers, we only need to resynchronize the decoding processes and continue the processing. Error resilience process contains four steps consisting of error detection, error localization, error recovery, and error concealment. The error detection is to detect errors in the transmission in order to make error recovery to prevent the decoding from being crashed. The error localization is to localize the exact positions of the corrupted MBs. Proper error localization help avoid the dropping of the subsequent and correct MBs, which can improve the overall video quality. Finally, error concealment can use the correlation of data that are correctly decoded to improve the video quality.

In the following, we first discuss the error detection, localization, and recovery algorithms for random bit error and packet loss, respectively.

2.1.2.1.1 Random Bit Error

Random bit error is occurred when there is a mismatch found during the decoding. The error is called as either a syntactic error or a semantic error. The instances of the syntactic error consist of the VOP start code error, the video packet start code (Resync_Marker) error and the VOP header error. The semantic error is detected when the following cases are found in decoding, which indicate that Marker_bit does not equal to 1, the stuffing_bit has wrong length, more than 64 coefficients are in a decoded block, the total amount of MBs in a VOP is larger than max MB_number, some illegal VLC symbols are parsed, the quantization scale is out of range, the motion vector is out of search range, next MB number is smaller than current MB number, etc. With the error detection techniques, we can find the error and avoid the decoding failure. For those undetected errors, a crash proof decoder is necessary to retain the video coding and service.

In [8] , the authors proposed a statistical method that can resume decoding processing. The basic idea is that the error may be found previously, we can back-track the decoding process for each MB and decide if the processing MB is wrong according to the statistical information. The statistical information used comes from the stationary properties of natural scene, the boundary across the corrupted blocks and the correct blocks will be very irregular and the statistical properties of a corrupted block vary from that of the reliable neighboring blocks.

2.1.2.1.2 Packet Loss

Packet loss is often detected with semantic errors as the occurrence of inconsistent MB numbers. When the MB numbers of nearby MBs are not consistent, there is a high probability that packets are lost over transmission channels. There are two conditions to detect the packet loss. One is that the decoding MB number is not monotonically increasing and the other is that the next MB number is smaller than the current MB number in the same frame. The latter indicates occurrence of at least one lost VOP header. For the general packet loss case, we can clearly recognize the MBs between current MB and next inconsistent MB are lost. When lost packets are detected, we can conceal all the lost MBs to recover the quality and continue decoding.

When any VOP headers get lost, the remaining MBs after the latest decoded MB in the current frame are all lost. After concealing the lost MBs, we still deal with the loss of VOP headers that contain critical information to decode a VOP. When the information is lost, we cannot decode the VOP correctly and thoroughly. We just could drop the remaining bits of the frame without VOP header until next VOP header is found. In case of VOP headers recognized as being lost, a whole frame is lost and frame level concealment is required. For frame level concealment, the direct duplication that is simple and has acceptable quality can be applied for error recovery. Other approaches with increasing complexity can be used to conceal the reconstructed frames when some VOP header is lost. The detailed description of the error concealment approaches will be given at the Chapter 3.

2.1.2.2 Error Resilience Tools

In the MPEG-4 Simple Profile specification, the error resilience tools consist of resynchronization marker (Resync_Marker), data partition and reversible variable length coding (RVLC). The error resilience tools can help resolve the occurrence of an error. The Resync_Marker is used to stop error propagation to the subsequent MBs of next packets. Data partition can be used to correct MVs for reconstructing MBs when any error is detected in the texture part. RVLC is used to localize error more exactly and can reduce the negative effect on the texture part. As to packet loss, the Resync_Marker can be used for detecting errors and resynchronize bitstream decoding. The three error-resilience tools provided by MPEG committee are summarized in the following paragraphs.

2.1.2.2.1 Resynchronization Markers

The resynchronization marker placed at the starting of a video packet is distinguishable from all possible VLC code words. Header information is also provided at the starting of a video packet following the resynchronization marker. This header contains the information necessary to restart the packet decoding process, which includes the first MB number in this packet and the quantization parameter (QP) necessary to decode the first MB. The MB number provides the necessary spatial resynchronization while the QP allows the differential decoding process to be resynchronized. Following the QP is the header extension code (HEC). As the name implies, the HEC is a single bit to indicate whether additional video object plane (VOP) level information will be available in the header. If the HEC is equal to one, then the following additional information (which is constant for each VOP and transmitted with the VOP header) is available in the packet header. The information consists of modulo time base, vop_time_increment, vop_coding_type, intra_dc_vlc_thr and vop_fcode_forward. The header extension feature enables the decoder to correctly utilize data contained in the current packet without reference to the packet containing the VOP headers. The header extension can also help error detection because it offers crosschecking capability since all packets in the same VOP share the same QP, time stamp, etc. When there is an error detected, we can find next Resync_Marker and resume the decoding process when next Resync_Marker is recognized.

There are other issues for error resilient streaming. As to packet length, a packet with smaller length contains less number of MBs, which may reduce the impact of any transmission error on the picture quality of received video. As for RTP header has minimal 12 bytes and resynchronization marker takes more than 2 bytes, each packet for streaming will takes more than 14 bytes. With the small packet length, increasing amount of overhead takes more bits to encode a video sequence, which decreases the transmission efficiency and bandwidth usage. Inversely, a packet with larger length can pack more MBs, which results in more quality loss when any transmission error occurs. Thus, for better quality of service, proper packet length is required for video streaming. The proper packet length is one of the important issues for optimizing media encoder and streamer.

2.1.2.2.2 Data Partitioning

Data partitioning can avoid discarding both texture and motion information of the MBs at the occurrence of any packet loss by separating the information into two parts. The separator is called as motion markers in Fig. 7 to Fig. 11. Data partition puts all the motion vectors of MBs in the packet together in the first part of a packet, inserts a motion marker, and then puts the residuals of MBs after the motion markers. When one of the two parts is corrupted, the information of the other part can be used for error recovery of the handling MBs. Motion information part is more important than the texture part in error recovery. When any error occurs in the texture residuals, we can use the correct motion vector to do the motion compensation and get the reconstructed MB. When an error occurs in the motion vector part, the correct residual may be used for error recovery. To model the effectiveness of using data partition, the authors in [26] have proposed 4 approaches for preventing errors.

(a) When an error is detected in the motion section, the decoder signals an error and losses all the MBs in the current packet

(b) When there is no error in the motion section and an error detected in the texture section, the decoded motion vectors are used to perform motion compensation.

(c) If no error is detected either in the motion or the texture sections but the Resync_Marker is not found at the end of decoding all the MBs in the current packet, an error is flagged and only the texture part is discarded.

(d) If no error exists in the motion part and the texture part and the Resync_Marker is found, but the MB number is inconsistent, we can rely on the current packet because of the number of MBs in the packet from the correct motion part.

In addition, since the DC components are important than the AC components in recovering the picture quality, the texture information is separated by a DC marker into DC and AC parts. With DC markers, we can reduce the impact of random bit errors in transmission channels on the quality of the received video.

Resync. Marker	MB. No.	QP	Combined Motion and DCT Data

Fig. 7. Traditional bit stream organization within the video packet [26] .

COD$_1$	MCBPC$_1$	CBPY$_1$	DQUANT$_1$	Encoded MV(s)$_1$	DCT data$_1$	……

Fig. 8. Bit stream components of each macroblock within the video packet [26] .

Resync. Marker	MB. No.	QP	Motion Data	MBM	DCT Data

Fig. 9. Bit stream syntax with data partitioning for motion and texture data [26] .

COD$_1$	MCBPC$_1$	Encoded MV(s)$_1$	COD$_2$	MCBPC$_2$	Encoded MV(s)$_2$	……

Fig. 10. Bitstream components of motion data [26] .

CBPY$_1$	DQUANT$_1$	CBPY$_2$	DQUANT$_2$	……	DCT Data$_1$	DCT Data$_2$	……

Fig. 11. Bitstream components of DCT data [26] .

2.1.2.2.3 Reversible Variable Length Coding (RVLC)

Reversible variable length codes (RVLC) are designed to decode the VLC symbols both in forward and backward. When an error is detected in the bitstream, we switch the symbol decoding from forward direction to backward direction to extract the correct texture data from the end of each packet, which can decrease the number of discarded bits and achieves good error localization. With data partition, RVLC in MPEG-4 is applied only to TCOEF coding used in texture part of each packet as shown in Fig. 12. Existing research works focus on generating a better RVLC and the shorter length. The process when the error is occurred in the RVLC is less discussed. In addition, for example, the MPEG-4 reference code with RVLC implementation modified by KDDI is more complicated. Therefore, more investigations are required for addressing the issues of optimized RVLC in performance optimization for error recovery and the complexity reduction for practical usage.

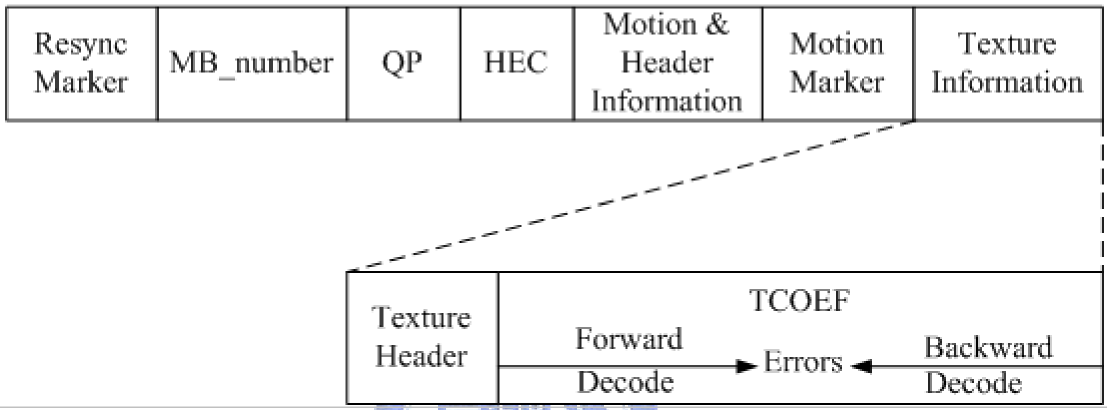


Fig. 12. Reversible variable length coding structure.

2.1.3 Summary

For better video service over the Internet or wireless channels, transmission error as bits error or packet loss shall be addressed. To address the better error recovery, MPEG-4 video coding standard provides three error resilience tools. Use of resynchronization makers can deal with the bits error or packet loss by resuming the decoding process from next resynchronization marker recognized. With the resynchronization markers, data partitioning and RVLC can handle random bit error to avoid the discarding of correct motion or texture information and localize the positions of wrong data precisely. With error resilience tools provided by MPEG-4 specification, some error concealment can improve the error recovery to provide video content of better quality.

2.2 Overview of Error Concealment Techniques

After detecting and localizing the lost MBs in a frame, we can use concealment methods to recover the damaged frame to a proper quality. With error concealment, the error in the current frame will not be propagated to the subsequent frames or MBs via the spatial and temporal prediction loops defined in the video compression standard.

Error concealment techniques can be basically classified into three categories including forward error concealment, error concealment by post-processing and interactive error concealment. The forward error concealment refers to the techniques in which the encoder adds redundancy to the bitstream to enhance the error resilience at the decoding process. The existing forward error concealment includes adaptive intra refresh [23] or forward error control (FEC) coding [27] . Error concealment by post-processing refers to decoding operations that can deal with the MBs recognized as being corrupted by transmission errors [20] . Error concealment by post-processing is a major research direction. In interactive error concealment, the transmitter and receiver cooperate to minimize the effect of transmission errors. The interactive methods heavily utilize feedback information provided by the receivers. Error concealment by post-processing can also be referred to as passive error concealment whereas the other two categories represent forms of active error concealment.

In addition, based on the processing domain, error concealment methods can be classified into spatial methods and temporal methods. Some hybrid approaches that combine both methods adaptively to fit video characteristics can get better performance. Based on considerations on realization complexity and computation cost, error concealment methods that are simple and can provide proper reconstruction quality are demanded in the decoder for practical usage.

2.2.1 Spatial Error Concealment

Spatial error concealment (SEC) techniques exploit the spatial redundancy within a picture. SEC techniques are mainly devoted to conceal I-pictures that have no motion information to be used for error concealment. For example, SEC techniques are applied to the first I-frame of a sequence or I-frames after scene changes. Considering the spatial correlation, the SEC approaches can provide better reconstructed video quality for the smooth or regular image blocks. For the irregular image blocks, SEC will produce the blurred image blocks, which is undesirable for subjective quality. The following paragraphs will introduce various SEC approaches and the impact on the MPEG-4 video.

A. Weighted Interpolation

Weighted interpolation interpolates each pixel of the entire 2*N**2*N* erroneous MBs with the adjacent pixels of the four adjacent MBs [6] . The interpolation operation is defined by

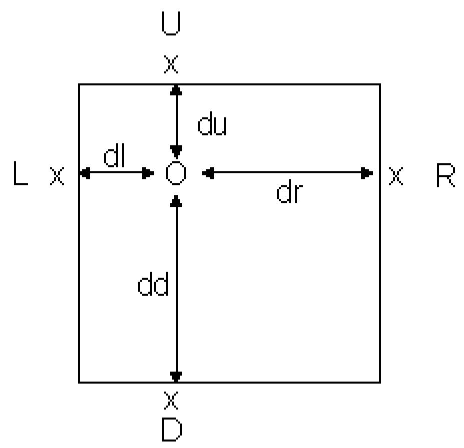


Fig. 13. Illustration of the weighted interpolation.

$$value = \frac{U * dd + D * du + L * dr + R * dl}{dd + du + dr + dl},$$

where *U*, *D*, *L* and *R* stand for the nearest correct pixels from the pixel that is concealed in the directions from the top, down, left and right directions respectively. The *du*, *dd*, *dl*, and *dr* are the distances between the current pixel and the pixel in the associated direction respectively.

When some of the MBs do not exist during interpolation, the distances at the related interpolation directions will be set to zero. In addition, weighted interpolation techniques have the best performance when all surrounding MBs exist. For existing video coding standards covering H.263 and MPEG-4, the casual coding scheme provides only the top and the left MBs for weighted interpolation.

Usually, with this technique, no abrupt transition will occur in an MB. However, the details and edges of the MB become blurred, which may result in visible blocking artifact between the adjacent blocks.

B. Quadrilinear Border Interpolation

In [4] , the authors interpolate the value of a missing pixel from the closest top, left, bottom and right closest pixels surrounding the handling pixel. Quadrilinear border interpolation is a little different from general methods that use MB based interpolation. In addition, to improve the quality, the methods need some refinements. Some modify the weights for the interpolation, which increases weighting to the pixel at the direction that has less distance. Some uses the average values of the nearest pixels and the two neighbors instead of the nearest pixels for the interpolation.

C. Other Interpolation Methods

To preserve the detail of the missing blocks, some other interpolation methods covering multi-direction interpolation [17] and adaptive recursive interpolation algorithm [25] have been proposed. To apply the detail preservation methods to the MPEG-4 video, some refinement to fit the casual coding procedure is required. For example, multi-directional filtering shall be reduced to the limited directions, which decreases the interpolation efficiency in reproducing the spatial details for error recovery.

D. Comparisons

Table 1 shows the comparisons of spatial concealment methods. Among all spatial concealment methods, we can see that bi-linear interpolation has medium complexity and high performance.

Table 1. Comparison of spatial concealment

Methods	Strength	Weakness
Spatial copy	Simple	Poor quality
Mean value	Simple	Blocking effect
Bi-linear interpolation	Good for smooth region	Less edge reservation
Multi-directional interpolation	Good for edge reservation	High complexity and high memory access
Quadri-linear border interpolation	Good for reconstructed pixels refinement	Not MB-based

2.2.2 Temporal Error Concealment

Temporal error concealment techniques exploit the temporal similarity of frames in a sequence. Temporal approaches are usually applied to P-pictures and B-pictures that have motion vectors. For a slow motion sequence, the motion information for the collocated MBs at two adjacent pictures is highly correlated, which indicates the motion vectors of the previous reference picture can also be used at concealment of the immediately following I-pictures. In addition, inconsistent motion vectors of the corrupted MB and the surrounding MBs may decrease the concealment efficiency of the temporal error concealment methods. The temporal algorithms are summarized in the following.

A. Direct Copy/Duplication

When the current blocks are lost, the simplest way is to conceal the current blocks by copying the block data at the same positions of the reference frame [6] . The direct copy method needs no additional memory, but suffers from visible shifts in the reconstructed picture when motion is present.

B. Motion Compensated Temporal Error Concealment (MC-TEC)

To improve the visual quality of direct copy approach, MC-TEC considers the motion information for error concealment. Based on the utilization of temporal correlation to compress the video sequence, motion vectors are introduced in video compression standard. For the lost MBs, we can assume that the motion information of the neighboring MBs has a close relationship. Based the relationship, we can use some motion vectors in the motion compensation to rebuild the lost MBs. The remaining issue is to derive the motion vectors of lost MBs. We can simply derive motion vectors of handling MB as the mean or median value of the motion vectors from the surrounding MBs [6] . Another way to choose a motion vector is to select a motion vector in a set of candidate motion vectors that have the minimum mean square error (MMSE) of the border pixels [6] . The method is called as spatial BMA(boundary matching algorithm) , as is shown in Fig. 14.

In [25], the authors have proposed an overlapped region approach for the recovery of lost motion vectors. The overlapped region approach can retain edge or signal continuity between the missing block and the neighboring blocks. In addition, the motion vector is re-estimated without any differential information from the neighboring blocks, which is called as temporal BMA. Fig. 15 shows the diagram of the overlapped region approach.

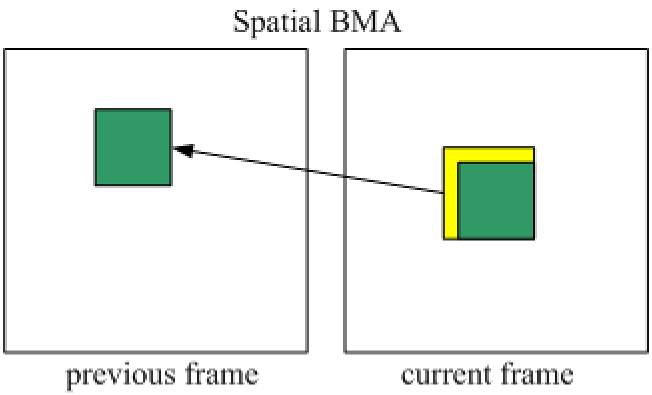


Fig. 14. Spatial BMA

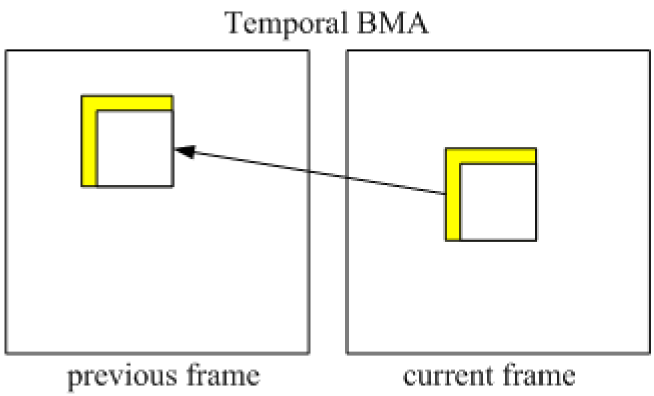


Fig. 15. Temporal BMA

In [4] , the authors have introduced two temporal concealment methods covering concealment by selection and concealment by search. Concealment by selection rebuilds the lost MBs based on the NEighbors MAtching criterion (NEMA). For each candidate motion vector, an error function is evaluated between the surrounding pixels and the motion compensated pixels of the extended area. Concealment by search is also called as Dynamic Motion Vector Estimation (DMVE). Each candidate motion vector is refined by examining all motion vectors in a specified search window surrounding the current MB under a NEMA-like criterion. DMVE can retain better recovered video with much more complexity as compared to the performance of the concealment by selection.

C. Comparisons

Table 2 shows the comparisons of temporal concealment methods. We can see that temporal BMA has medium complexity and high performance.

Table 2. Comparison of temporal concealment

Methods	Strength	Weakness
Zero motion	Simple	Poor quality in motion part
Candidate MV compensation	Simple	Poor quality by blind choice
Spatial BMA	Less blocky effect	Poor quality when edges are on the boundary
Temporal BMA	Good edge preservation on the boundary	Poor quality with less temporal correlation
DMVE	Better reconstructed quality with more MVs	High complexity due to more MV tested and more surrounding lines

2.2.3 Hybrid Error Concealment

For better error concealment, a straightforward concept is to adapt the best concealment technique based on video properties, which is called as hybrid error concealment. At frame level, we can use spatial error concealment for I-frames and temporal error concealment for P-frames and B-frames. At MB level, spatial and temporal concealment are adopted based on the properties of the handling frame. The remaining issue of realizing optimized hybrid error concealment schemes is to derive the criterion of choosing the best concealment technique for each MB.

The selection criteria of the best concealment method are based on the measurement of the image activities including spatial activities and temporal activities.

In [6] , the authors have provided that the amount of the temporal activities is derived from the mean square error (MSE) between two MBs of adjacent pictures. The spatial activities can be computed by measuring the variance (VAR) of the nearest neighboring MB (top or bottom) of the current picture. When the amount of temporal activities in MSE is smaller than a certain threshold T1=VAR, the temporal concealment is applied. Otherwise, the spatial concealment is used.

In [24] , the authors have proposed another decision method based on prediction error statistics measured in the neighborhood. The decision region is shown in Fig. 16. The VAR is computed by

$$VAR = E[(X - \overline{X})^2], VAROR = E[X^2] - \mu^2,$$

where X is the neighboring good MB data, $\overline{X}$ is the data of the corresponding MB in the previously decoded frame at the co-located position and μ is the average value of the neighboring good MB in the current frame. One can appreciate that VAR may indicate the local motion and VAROR can present the local spatial details. If VAR > VAROR, the better concealment method is spatial interpolation. When VAR$\leq$VAROR or VAR$\leq T$, where T is a preset threshold, the better concealment method is temporal replacement.

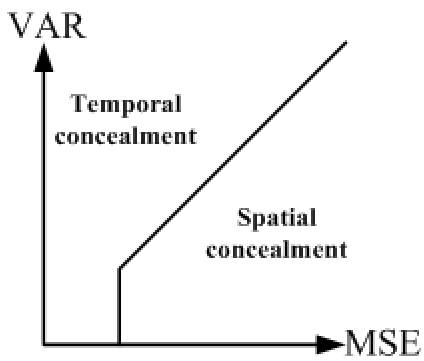


Fig. 16. Adaptive error concealment strategy

In [8] , the authors have given a set of concealed block candidates for the corrupted block and proposed a fitness function for error concealment to select the “best” concealed block candidate among the set. The set includes several types of concealed block candidates covering neighboring blocks, average or median of neighboring blocks, motion compensation blocks with motion vectors from neighboring blocks, average or median of motion compensation blocks with motion vectors from neighboring blocks…etc. To select the best concealed block, some parameters are calculated based on the following information consisting of smooth boundary, similar statistical properties, and temporal similarity with the motion compensated block in the previous picture. Finally, with the parameters, a fitness function is used to decide the best candidate for concealing each block.

In [9] , the authors have adopted boundary matching errors (BME) for various combinations of MBs. The MB combinations cover the top and top-left, the top and top-right, the top and bottom-left, top and bottom-right, the temporal concealed and top, and the temporal concealed and bottom. With MBE values, the smoothness and similar property between concealed MB and neighboring MBs can be verified. The boundaries which are used are shown as Fig. 17.

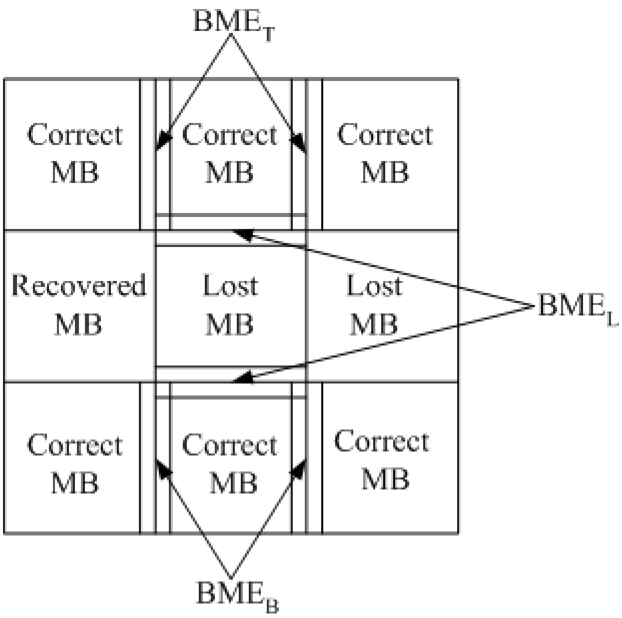


Fig. 17. Boundary smoothness property method

In [4] , some strategies using temporal concealment in I-VOP or spatial concealment in P-VOP are discussed. At the scene change or fading pictures, the spatial redundancy may be more than the temporal redundancy. The majority of MBs in I-VOPs are referred as INTRAs. When the INTRAs have some temporal redundancy between the temporally successive VOPs, temporal methods will achieve better visual quality than the spatial algorithms.

2.2.4 Intra Refreshment

To enhance the robustness of delivered video contents, R-D optimized intra-refresh (RDIR) has been proposed for solving error propagation more effectively [25] . Intra refresh technique inserts intra blocks instead of inter blocks in the predicted frames (P-frames) to prevent error propagation over error-prone network. Since the intra block uses more bits, it is inefficient when the network condition and object motion vary over time. To resolve this problem, intra block insertion with adaptive to channel condition and motion information can retain the R-D optimization performance of an encoder with intra refreshment.

In RDIR [25] , for each macroblock (MB), we compute the cost for intra and inter blocks said J_{intra} and J_{inter} by the Lagrangian formula $J = D + \lambda \cdot R$, where J is the Lagrangian cost and λ is the parameter used to control coding bit rate in encoding process. The symbol D means the distortion induced from residual quantization and R equals to the number of bits used in coding a MB.

To quantify the distortion D, coding characteristics of each coding mode in the specification of MPEG-4 Simple Profile are considered. For inter mode, each MB has the concealment distortion coming from the reference frame and the quantization distortion of the current MB. Thus, for the handling MB $D = D_{inter}$ is defined as the weighting summation of the concealment distortion and the quantization distortion.

$$D_{inter} = D_q \times (1 - p) + D_c \times p \qquad (1)$$

Where D_q is the quantization error and D_c is the concealment error from previous frame. The symbol p denotes the channel packet loss rate. For intra mode, since there is no concealment error from the previous frame for the current MB, the distortion is set as $D_{inra} = D_q$.

To achieve the R-D optimization under the proposed intra-refresh encoding, the parameter λ is updated every frame to control the bits used under the same distortion. In addition, RDIR increases overall bitrate or decreases the overall visual quality due to the use of more intra blocks. Thus, the prediction residual for each MB within P-frames is selectively transmitted. The MB, which has the energy of prediction residual less than a specified threshold, will be zeroed out and is transmitted as a SKIP MB defined in MPEG-4 video standard. Therefore, the proposed coding system using the improved RDIR can retain almost identical bitrate with insignificant video quality degradation. In addition, we adopted direct copy scheme to conceal error and quantify the distortion D_c at the encoder.

2.2.5 Summary

The general error concealment methods are divided into two types including spatial error concealment and temporal error concealment. Spatial methods based on spatial correlation recover the picture quality. For spatial concealment, interpolation is simple, easy to implement, and has good performance. Temporal concealment using temporal correlation recovers the visual quality. The basic methods for temporal concealment use some candidate motion vectors and some cost functions to decide a best motion vector. In addition, to have a better performance based on the video characteristics, some hybrid methods are investigated. To further avoid the blocking artifacts, the smoothness of the concealed video frames is judged with boundary matching or statistical similarity.

2.3 Overview of MPEG-21 Standard

MPEG-21 is an open standards-based framework for multimedia delivery and consumption [28] , [29] . The MPEG-21 vision is to define a multimedia framework to enable transparent and augmented use of multimedia resources across a wide range of networks and devices used by the different communities. The concerns of MPEG-21 are: management of content, repurposing content based on user preferences and device capabilities, protection of rights, protection from unauthorized access/modification, protection of privacy of providers and consumers, and so on. Universal Multimedia Access (UMA) deals with delivery of audio, images, video, and multimedia content under different network conditions, user and publisher preferences, and capabilities of terminal devices. The primary motivation of UMA is to enable terminals with limited communication, processing, storage and display capabilities to access rich multimedia content.

The basic concepts in MPEG-21 relate to what and who within the multimedia framework. What is a Digital Item that’s a structured digital object with a standard representation, identification, and metadata within the MPEG-21 framework. The who is a user who interacts in the MPEG-21 environment or uses a Digital Item, including individuals, consumers, communities, organizations, corporations, consortia, governments and other standards bodies, and initiatives around the world. The user roles include creators, consumers, rights holders, content providers, distributors, and so on—there's no technical distinction between providers and consumers. All parties that must interact within MPEG-21 are categorized equally as users. They assume specific rights and responsibilities according to their interaction with other users. All users must also express and manage their interests in Digital Items.

In practice, a Digital Item is a combination of resources, metadata, and structure. The resources are the individual assets or (distributed) content. The metadata describes (distributed) data about or pertaining to the Digital Item as a whole or also to the individual resources in the Digital Item. Finally, the structure relates to the relationships among the parts of the Digital Item—both resources and metadata. An example of a Digital Item might be a presentation of a university, including photos, videos, animation graphics, textual information, news related to the university's research activities, e-learning material, navigational information driven by user preferences, and so on. The Digital Item is thus the fundamental unit for distribution and transaction within the MPEG-21 framework.

Although electronic content creation, distribution, consumption, and trade are already possible today, no current end-to-end solutions let different user communities interact in an interoperable and efficient way. MPEG-21 guarantees such interoperability by focusing on how the elements of a multimedia application infrastructure should relate, integrate, and interact. The strength of MPEG-21 is that it's a standard grown from clear multimedia industry requirements for interoperability. As such, the requirements, existing standards, and current work are firmly based on usage cases and scenarios provided by that industry.

2.3.1 Overview of MPEG-21 Part-7 Digital Item Adaptation

The goal of the Terminals and Networks key element is to achieve interoperable transparent access to (distributed) advanced multimedia content by shielding users from network and terminal installation, management and implementation issues. This will enable the provision of network and terminal resources on demand to form user communities where multimedia content can be created and shared, always with the agreed/contracted quality, reliability and flexibility, allowing the multimedia applications to connect diverse sets of Users, such that the quality of the user experience will be guaranteed.

Towards this goal the adaptation of Digital Items is required [30] -[31] . This concept is illustrated in Fig. 18. As shown in this conceptual architecture, a Digital Item is subject to a resource adaptation engine, as well as a descriptor adaptation engine, which produce together the adapted Digital Item.

It is important to emphasize on that the adaptation engines themselves are non-normative tools of Digital Item Adaptation. However, descriptions and format-independent mechanisms that provide support for Digital Item Adaptation in terms of resource adaptation, descriptor adaptation, and/or Quality of Service management are within the scope of the requirements.

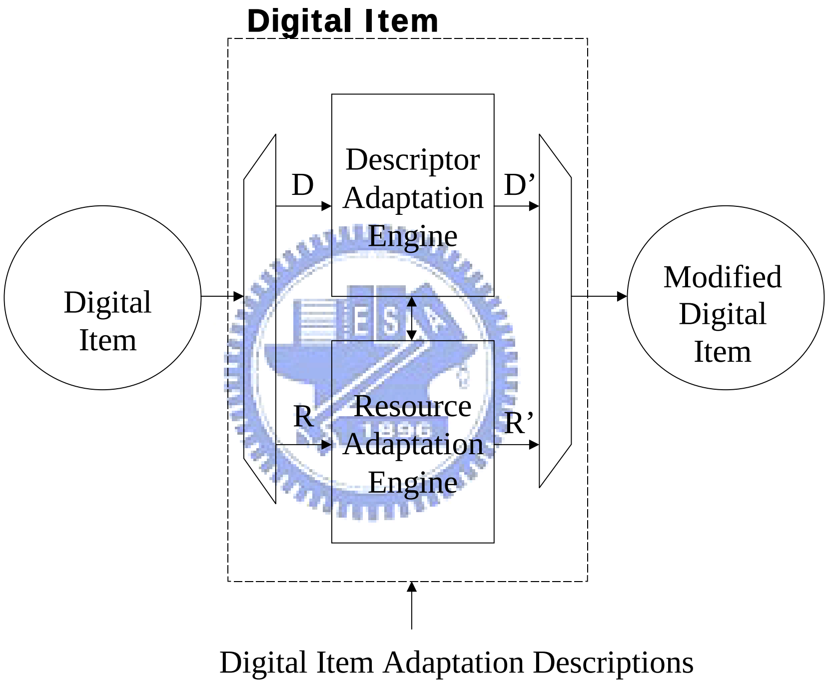


Fig. 18. Concept of MPEG-21 Part-7 Digital Item Adaptation.

In May 2002, a number of responses to the Call for Proposals on MPEG-21 Digital Item Adaptation were received. Based on the evaluation of these proposals, a Working Draft has been produced. The specific items targeted for standardization are outlined below.

- User Characteristics: Description tools that specify the characteristics of a User, including preferences to particular media resources, preferences regarding the presentation of media resources, and the mobility characteristics of a User. Additionally, description tools to support the accessibility of Digital Items to various users, including those with audio-visual impairments, are being considered.

- Terminal Capabilities: Description tools that specify the capability of terminals, including media resource encoding and decoding capability, hardware, software and system-related specifications, as well as communication protocols that are supported by the terminal.

- Network Characteristics: Description tools that specify the capabilities and conditions of a network, including bandwidth utilization, delay and error characteristics.

- Natural Environment Characteristics: Description tools that specify the location and time of a User in a given environment, as well as audio-visual characteristics of the natural environment, which may include auditory noise levels and illumination properties.

- Resource Adaptability: Tools to assist with the adaptation of resources including the adaptation of binary resources in a generic way and metadata adaptation. Additionally, tools that assist in making resource-complexity trade-offs and making associations between descriptions and resource characteristics for Quality of Service are targeted.

- Session Mobility: Tools that specify how to transfer the state of Digital Items from one User to another. Specifically, the capture, transfer and reconstruction of state information will be specified.

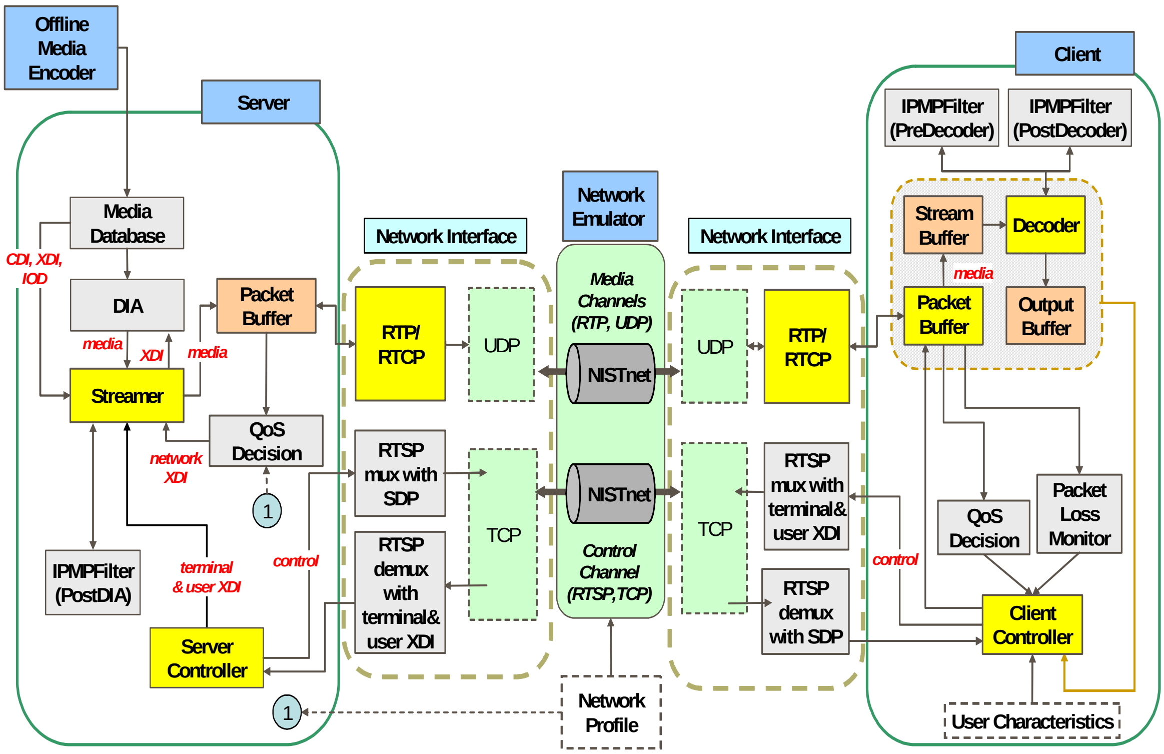


Fig. 19. System architecture of the NCTU FGS-based video streaming test bed [32]

2.3.2 Overview of MPEG-21 Part-12 Multimedia Test Bed

The goal of FGS-based streaming test bed is to support MPEG-21 DIA scheme with a more strict evaluation methodology according to the specified common conditions for scalable coding. Fig. 19 shows the NCTU test bed system architecture, which covers four key modules including the FGS-based Video Content Server, Video Clients, Network Interface, and Network Simulator. The detailed description of each module can be found at [32] .

2.3.3 Summary

MPEG-21 gives a provision to realize a digital life with various multimedia data by network or communications to deliver multimedia everywhere anytime. Digital item adaptation makes the digital items adaptive to the various service conditions covering terminal capability, network condition, user preference, etc. The multimedia test bed provides a platform to simulate real network conditions for video content delivery. With the digital item adaptation, multimedia test bed, and other parts of MPEG-21, we could realize universal multimedia access environment in a near future.

Chapter 3

Adaptive Spatial/Temporal Error Concealment

The streaming video system is built based on a subset of visual coding tools of MPEG-4 video coding standard. In addition, the streaming system provides error resilient decoder to serve better video service. Error resilience can recover the error occurred at transmission channels. To get better quality when any packet is lost, error concealment algorithms are employed. Finally, the streaming system is integrated into a multimedia test bed that can simulate a variety of network conditions and evaluate error resilience performance for on-line testing.

With a good error resilience system and good error concealment methods, we can provide a good video service on communications or content delivery on the Internet or over the wireless channels. The error resilience capabilities can detect and localize lost MBs. In addition, the error resilience tools can recover the decoding process from the transmission errors. The error concealment can improve the reconstructed video quality to serve video contents of high quality. To evaluate the error resilience capabilities and error concealment algorithms, we realize an MPEG Simple Profile decoder and embed the decoder into the MPEG-21 part-12 multimedia test bed in Fig. 19.

3.1 Error Resilient Decoding System

To provide good video streaming via the Internet and wireless channels, a decoding process against the packet loss is a major issue. When any packet loss occurs, we resume the decoding process and recover the reconstructed video. In addition, we realize a crash proof decoder to continue decoding of video bitstream.

3.1.1 Architecture of Error Resilient Decoder

The decoding architecture is as Fig. 20. The middle parts are original decoder blocks. The upper parts are error resilience blocks including error detection, error recovery, and error localization. Besides, timing check and correction for deal with VOP header loss is added. The lower parts are error concealment blocks including hybrid scheme, spatial and temporal concealment, and smooth filter. The modification of adding error resilience parts and error concealment parts can realize a robust decoding system.

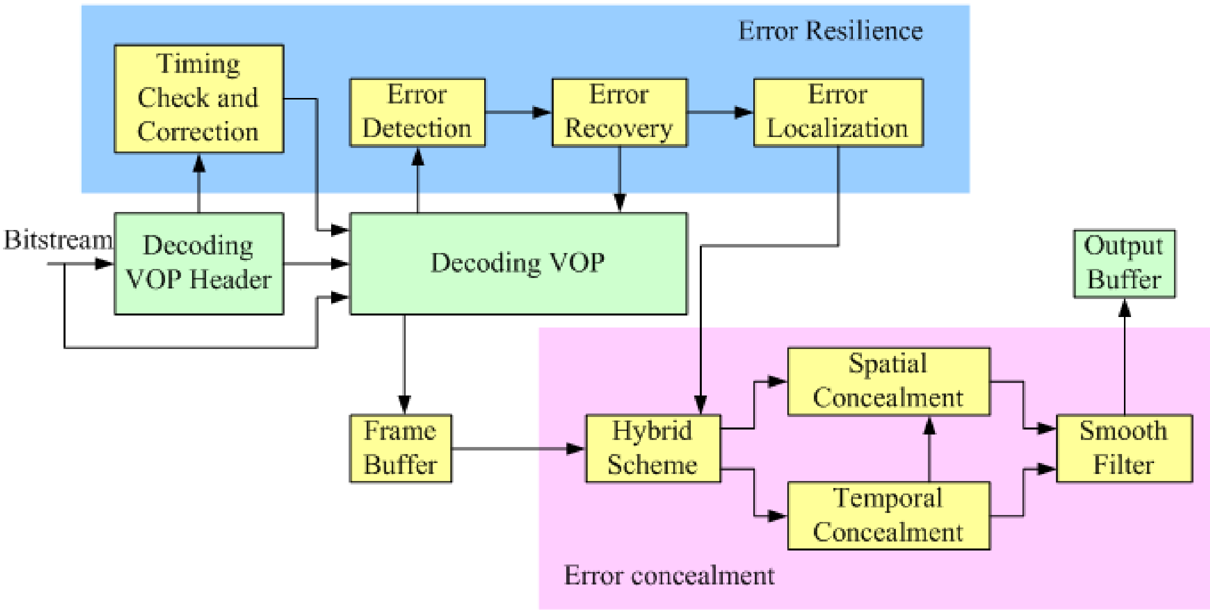


Fig. 20. Architecture of Error Resilient Decoder

3.1.2 Error Resilience for Packet Loss

In the decoding process, the packet loss is detected with the semantic consistency of MB (macroblock) number between the adjacent video packets. After decoding a video packet within the received bitstream, we search for the next resynchronization marker, which starts a new packet. At each packet, when the first MB number is not consistent with the MB number of last decoded packet, the packet loss is detected.

3.1.2.1 Packet Loss Without VOP Header Loss

After detection of packet loss, the MBs with the MB numbers between the two MB numbers successfully recognized are set as being lost. The lost MB numbers are record for error concealment when finishing decoding the VOP. After fully decoding a VOP (video object plane), the error concealment process will conceal the corrupted MBs.

3.1.2.2 VOP Header Loss

At picture level, any lost VOP header drops the first few MB data of the current VOP. Without the first MB data, the intra prediction or DC/AC prediction become incorrect that we need to discard the remaining MBs of the same VOP. To conceal the missing VOP, picture-level error concealment algorithms are used. In addition, time stamp in the VOP header is important for resuming decoding process and resynchronizing display time.

In an error resilient system, VOP-level error concealment is required to recover from the VOP header loss during transmission. When the VOP header loss, the error will be propagated to all MBs of the current VOP via the prediction loops including intra frame prediction and inter frame prediction. Thus, all data of the same VOP are discarded. To retain the same frame rate, the damaged VOP are rebuilt with direct copy/duplication methods that take less computation cost and memory.

In the reference software of MPEG-4 Simple Profile coding standards, the header information with the video packet headers is used to detect the occurrence of any VOP header loss. In MPEG-4 video specification, "*mb_a*" means the first MB number of current received packet and "*mb_b*" is the first MB number of next received packet. "*mbnum*" indicates the latest MB number that is validly reconstructed. The reference software detects the packet loss by two steps.

1. After decoding a packet, we read next packet header and get *mb_b*. As a valid packet is fully decoded, a counter is increased by one
2. If the difference between *mb_b* and *mb_a* is not equal to the counter, we conceal MBs that have MB numbers from *mbnum* to *mb_b*.

The VOP header error is one of the semantic errors. VOP header lost can be detected by checking if *mb_b* will smaller than *mb_a* or the current VOP display time is smaller than the previous VOP display time. With any VOP header loss, we conceal the lost VOPs via direct duplication to retain the temporal resolution of the receive video. In addition, we resume the decoding process by seeking to a new VOP start code.

3.2 Algorithm Descriptions

Based on the error resilient system, error concealment can use the localization of lost MBs to conceal the corrupted VOP. Good error concealment needs some consideration to make the performance well. In this sector we will introduce the concept of our error concealment. Three innovations are described to improve the concealment quality.

3.2.1 Rationale

Error concealment uses the localization of lost MBs and neighboring relevant data of lost MBs to conceal the corrupted VOP. To have good concealment results need to combine two key points. A simple and high performance method and use relevant data as much as possible will be described as follows.

3.2.1.1 Simple and High Performance Method

In this thesis, the spatial concealment reconstructs the pixels of each block based on linear interpolation of pixels at the neighboring blocks. In addition, the temporal concealment is referred as a block-based boundary matching algorithm (BMA), which is derived with the motion vectors and boundary pixels of each MB or each block. In [4] , the authors have provided a refinement to make the spatially concealed pixels smoother. The refinement method can cooperate with the post filter to reduce the blocky effect of the reconstructed VOPs. The combination of the refinement and the post filter will be explained at the following paragraph.

Prior to the application of error concealment, the criterion of boundary matching algorithm (BMA) is applied to calculate the variation of the boundary pixels within every block. By the derivation approach, the BMA can be further classified into two different methods including spatial BMA and temporal BMA. The spatial BMA checks the difference between the reconstructed block and the boundary pixels around the block. Temporal BMA checks the difference between the surrounding pixels in current frame and the surrounding pixels in reference frame. The spatial and temporal BMA are derived by $SAD_{spatial}$ and $SAD_{temporal}$ in sum of absolute difference (SAD), respectively.

$$SAD_{spatial} = \sum_{j=0}^{7} | P_k(x, y+j) - P_{k-1}(x-1, y+j) | + \sum_{j=0}^{7} | P_k(x+15, y+j) - P_{k-1}(x+16, y+j) | + \sum_{i=0}^{7} | P_k(x+i, y) - P_{k-1}(x+i, y-1) | + \sum_{i=0}^{7} | P_k(x+i, y+15) - P_{k-1}(x+i, y+16) | \quad \textbf{(1)}$$

$$SAD_{temporal} = \sum_{j=0}^{7} | P_k(x-1, y+j) - P_{k-1}(x-1, y+j) | + \sum_{j=0}^{7} | P_k(x+16, y+j) - P_{k-1}(x+16, y+j) | + \sum_{i=0}^{7} | P_k(x+i, y-1) - P_{k-1}(x+i, y-1) | + \sum_{i=0}^{7} | P_k(x+i, y+16) - P_{k-1}(x+i, y+16) | \quad \textbf{(2)}$$

The valuation function to check the boundary matching can be mean square error (MSE). According to our observations, the performance by SAD is almost identical to that by MSE. In addition, SAD has much less computation complexity. Consequently, we use SAD as our evaluation function to seek for minimum BMA. The remaining issue of the proper evaluation function for spatial or temporal BMA needs further investigation.

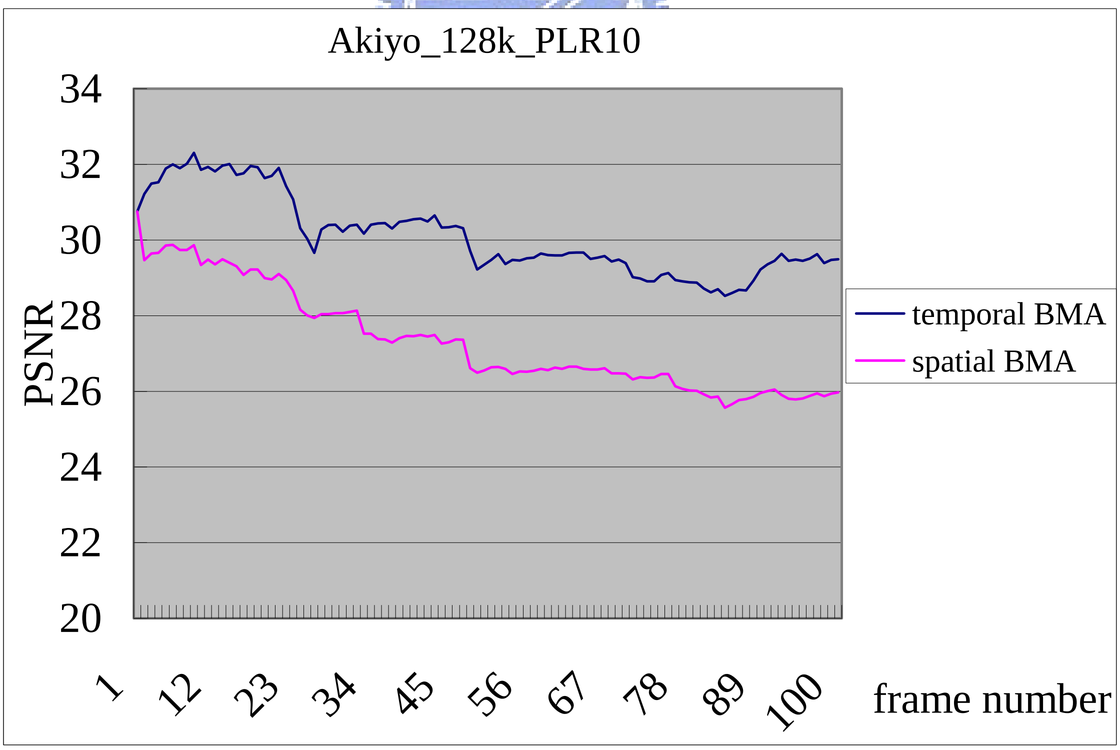


Fig. 21. PSNR of Akiyo 128k bits/sec with packet loss rate 10%

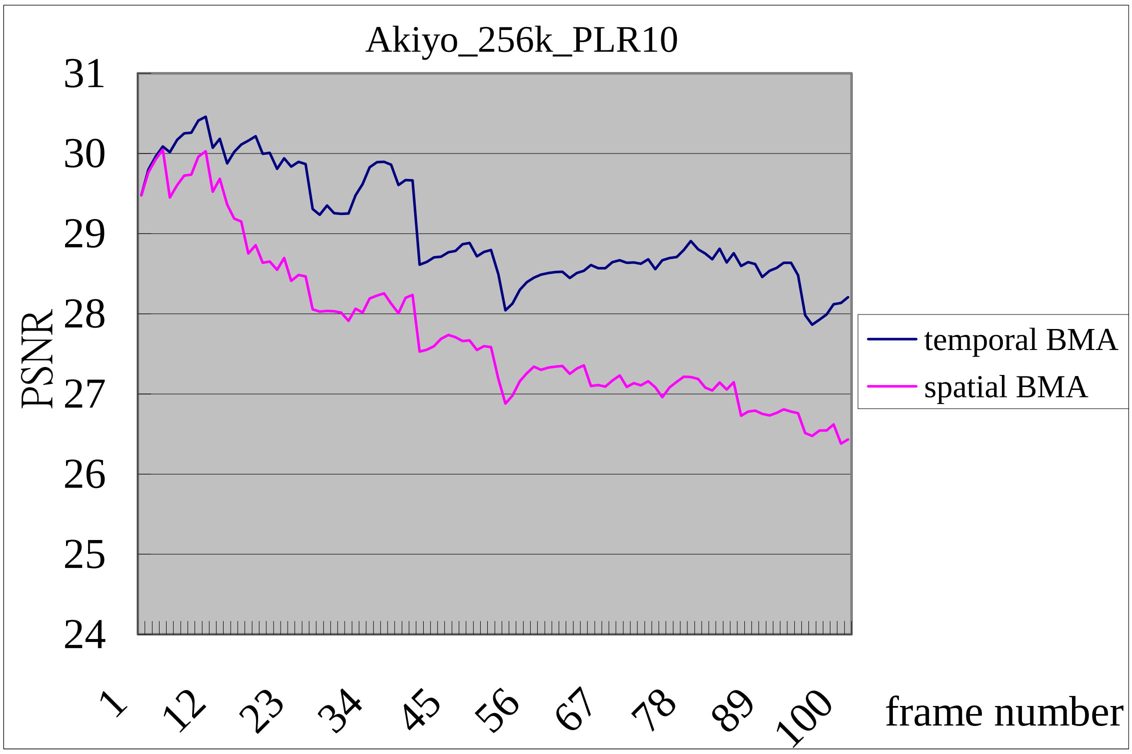


Fig. 22. PSNR of Akiyo 256k bits/sec with packet loss rate 10%

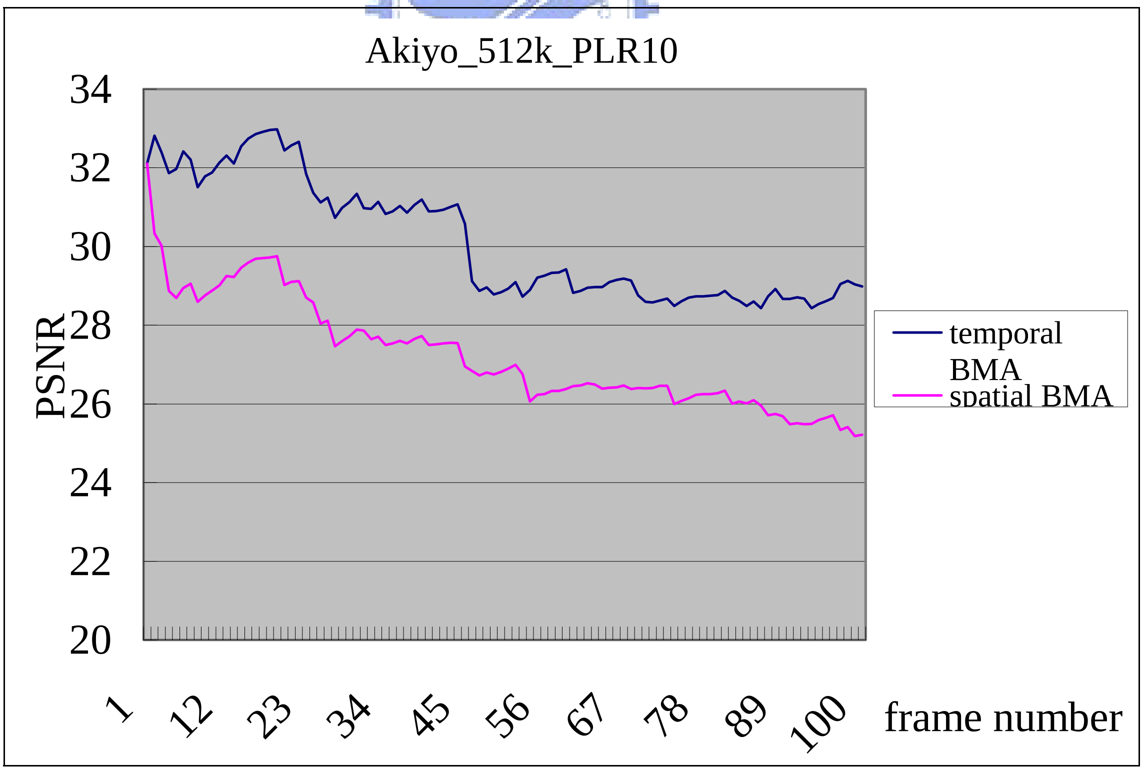


Fig. 23. PSNR of Akiyo 512k bits/sec with packet loss rate 10%

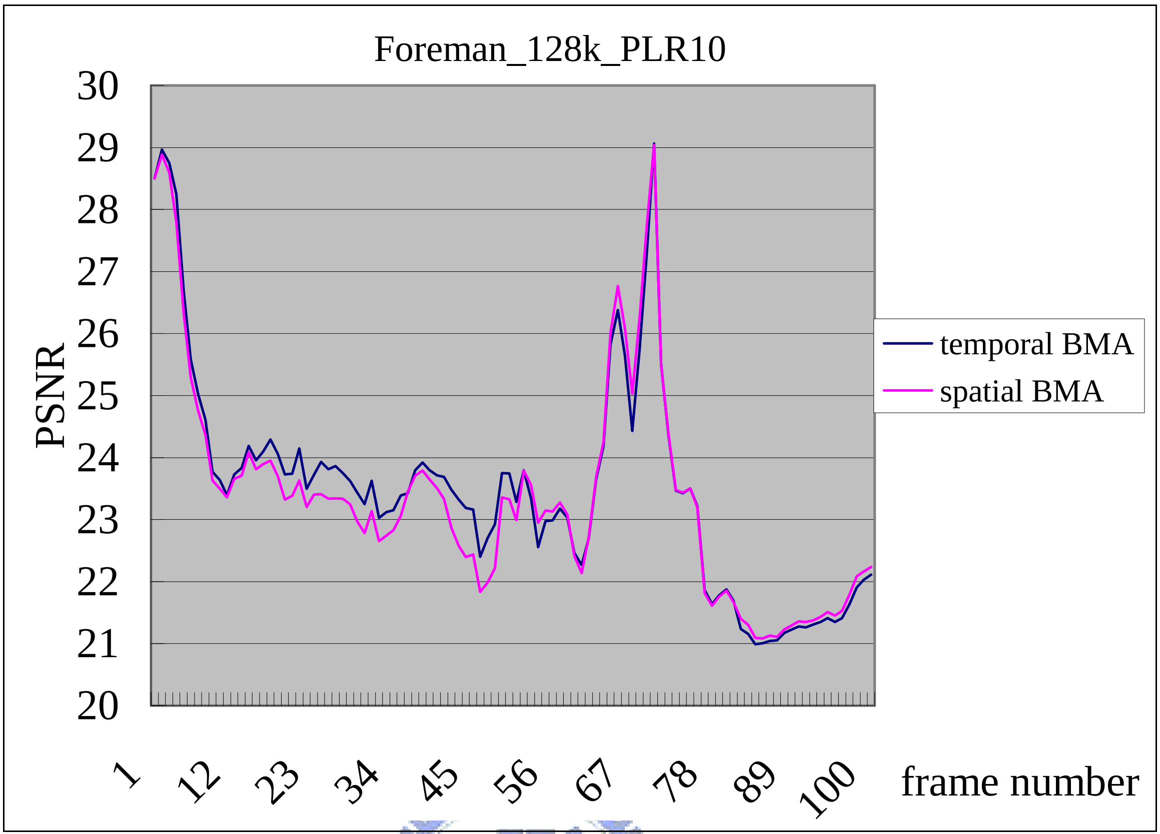


Fig. 24. PSNR of Foreman 128k bits/sec with packet loss rate 10%

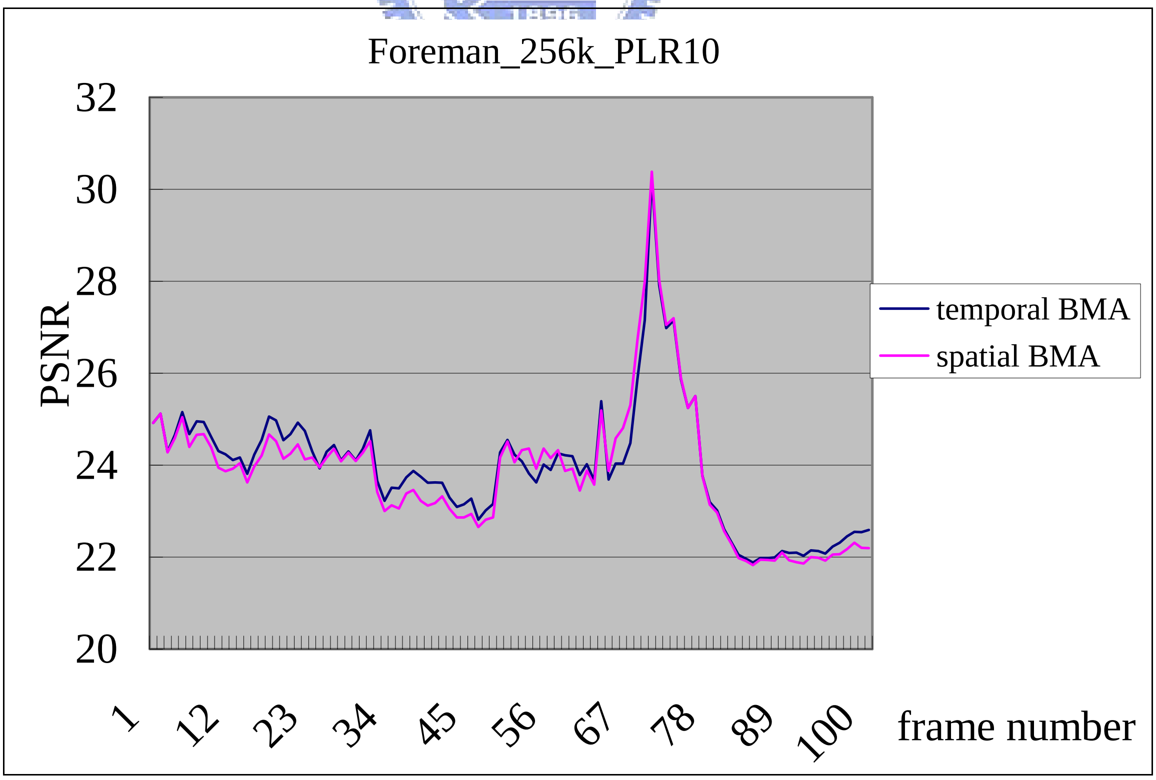


Fig. 25. PSNR of Foreman 128k bits/sec with packet loss rate 10%

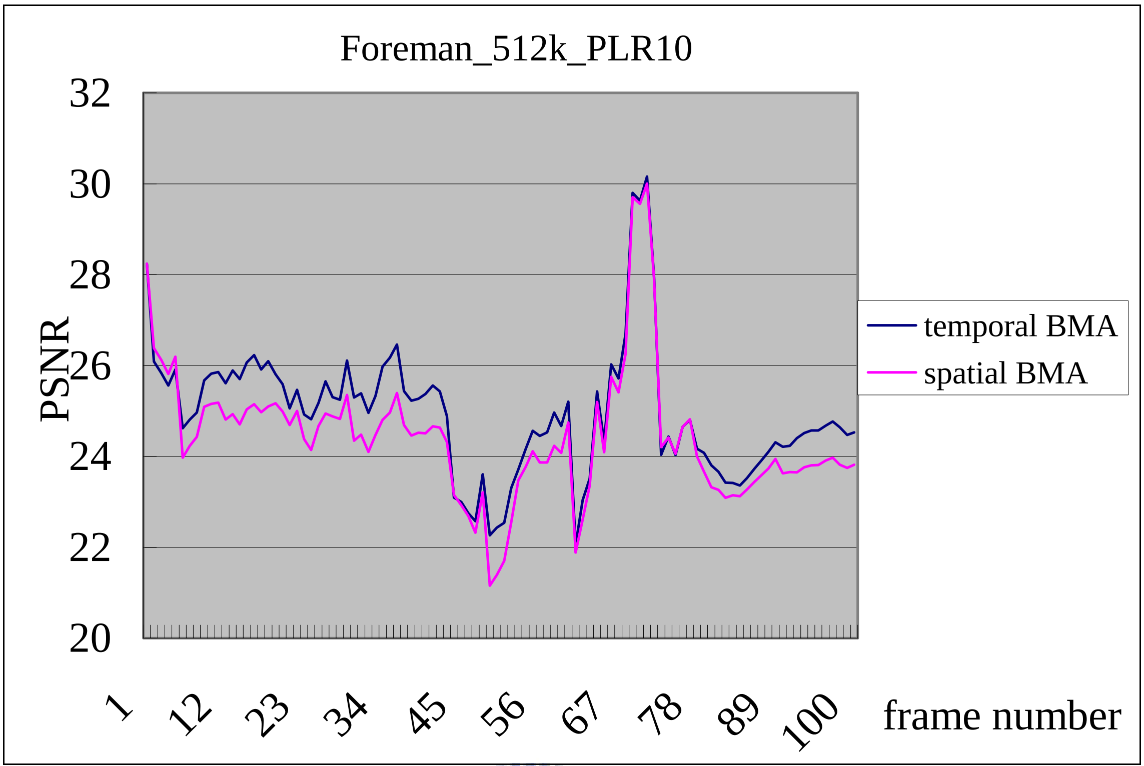


Fig. 26. PSNR of Foreman 128k bits/sec with packet loss rate 10%

3.2.1.2 Use Relevant Data as Much as Possible

All the concealment methods use surrounding relevant information to recover the corrupted MBs. When the relevant data are fewer, the concealment result is poorer. Using more relevant data can make error concealment methods work well. Taking the factor into consideration, we conceal lost packet in bi-direction as Fig. 27.

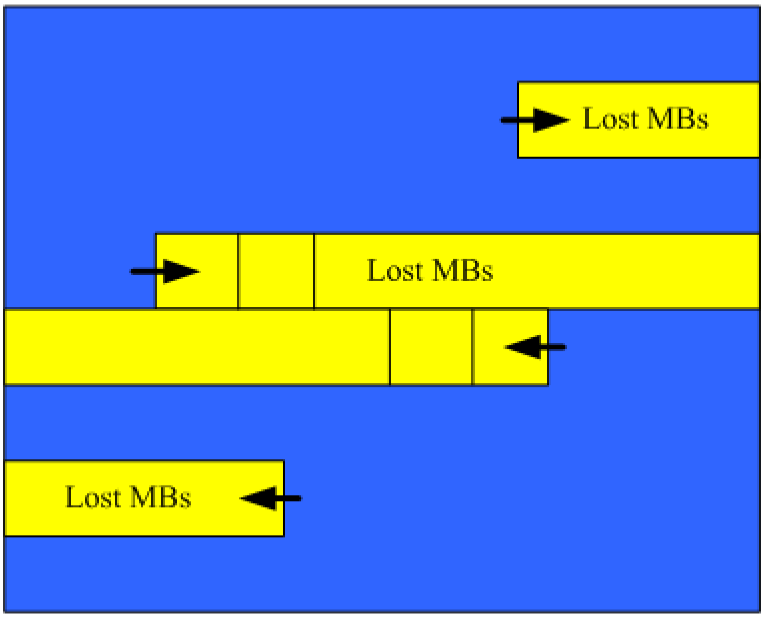


Fig. 27. Bi-directional concealment.

3.2.2 Innovations

There are three innovations in our algorithm. The first is using a less complexity hybrid scheme to choose when to use spatial concealment or temporal concealment. The second one is to implement block-based concealment to refine general MB-based method. Finally, a simple smoothing filter is used for improving visual quality.

3.2.2.1 Hybrid Scheme

The application of spatial or temporal concealment methods is decided based on the spatial characteristics and temporal similarity of received video. For spatial concealment, interpolation using surrounding reliable pixels is a simple and widely used to recover the lost packets of the I-frame. For temporal concealment, boundary matching algorithm (BMA) [2] is used to re-estimate the best motion vectors for motion compensation of lost MBs. Based on the observations, spatial concealment can fit fast motion or low detailed sequences since the correlation across successive frames is smaller than the correlation of pixels within the frame. In other words, temporal concealment is suitable for slow motion or highly detailed sequences. The temporal concealment can avoid visible blocky artifact introduced by the spatial concealment. Thus, an adaptive temporal/spatial error concealment scheme is present to provide video contents of better picture quality.

Several considerations to select spatial error concealment or temporal error concealment and block-based concealment consideration are included in our proposed adaptive hybrid error concealment method.

3.2.2.1.1 Inferior result by temporal concealment

Reference hybrid concealment methods use some statistics characteristics such as temporal activity, spatial activity, or boundary similarity to decide to use spatial concealment or temporal concealment. The methods take more extra computational complexity to get the information. We take the result of BMA to decide whether the result is good enough or not. If the boundary difference from BMA result is larger than the threshold, we use spatial concealment to conceal the MB which may have less temporal correlation. If the boundary difference from BMA result is smaller than the threshold, we just use the result of temporal concealment to conceal the MB.

3.2.2.1.2 Fast Motion

In the fast motion area or scene change, the temporal correlation may become very low and motion vectors will be in great confusion or intra blocks are added. By observing the motion vectors in the sequence, when the motion vector is large, the correlation between surrounding motion vectors are very low because of fast motion or motion in great confusion. We will use spatial error concealment when detect large motion vectors.

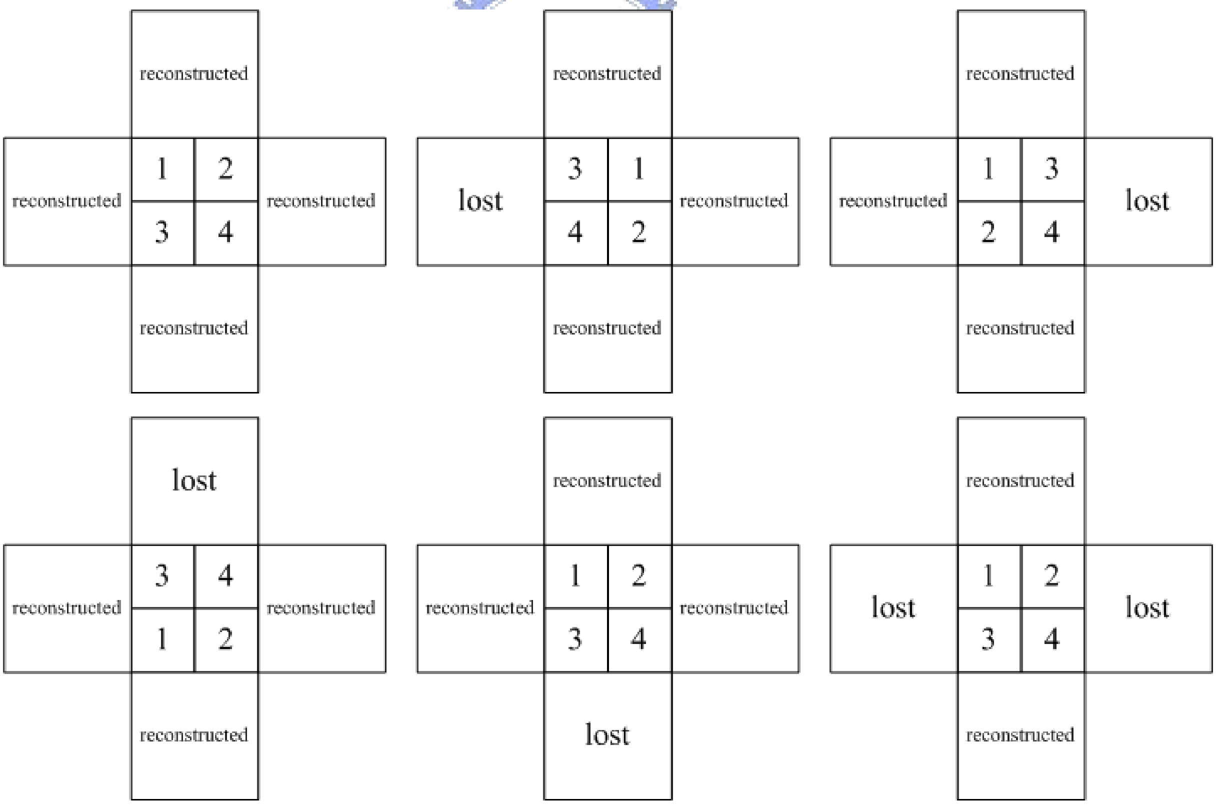
3.2.2.1.3 Surrounding Intra

In the fast motion area or scene change, the temporal correlation may become very low and motion vectors will be in great confusion or intra blocks are added. When the intra blocks are more, the surrounding motion vectors are less and we can’t have enough temporal correlation to recover the MB. We will use spatial error concealment to conceal the MB.

3.2.2.2 Block-Based Refinement

Considering the strong correlation of pixels within a small area and fit the 4-MV coding mode used by MPEG-4 Simple Profile, the block-based error concealment adopts an 8x8 block as a processing unit. Based on validation of 4 surrounding MBs and the location of the current block, each of 4 8x8 blocks can be concealed in different orders. For example, according to the validation of the 4 neighbors, there are 15 conditions of concealment order as shown in Fig. 28. The numbers within the central MB indicate the concealment order of a MB. The block-based refinement can let both spatial and temporal concealment in a single MB.

Combined with hybrid concealment scheme and block-based refinement, we can have an error concealment flow as Fig. 29.



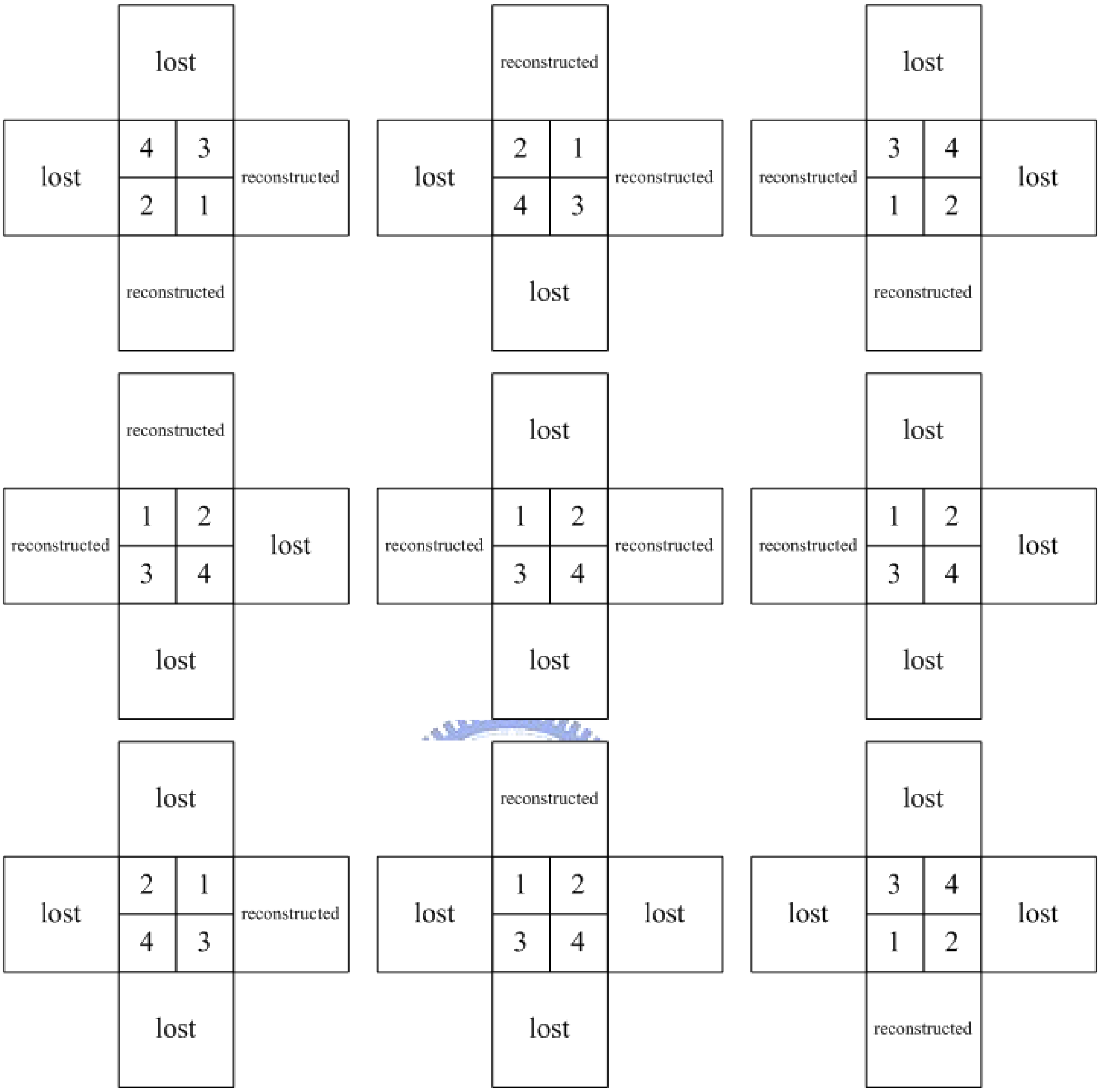


Fig. 28. Order of block concealment.

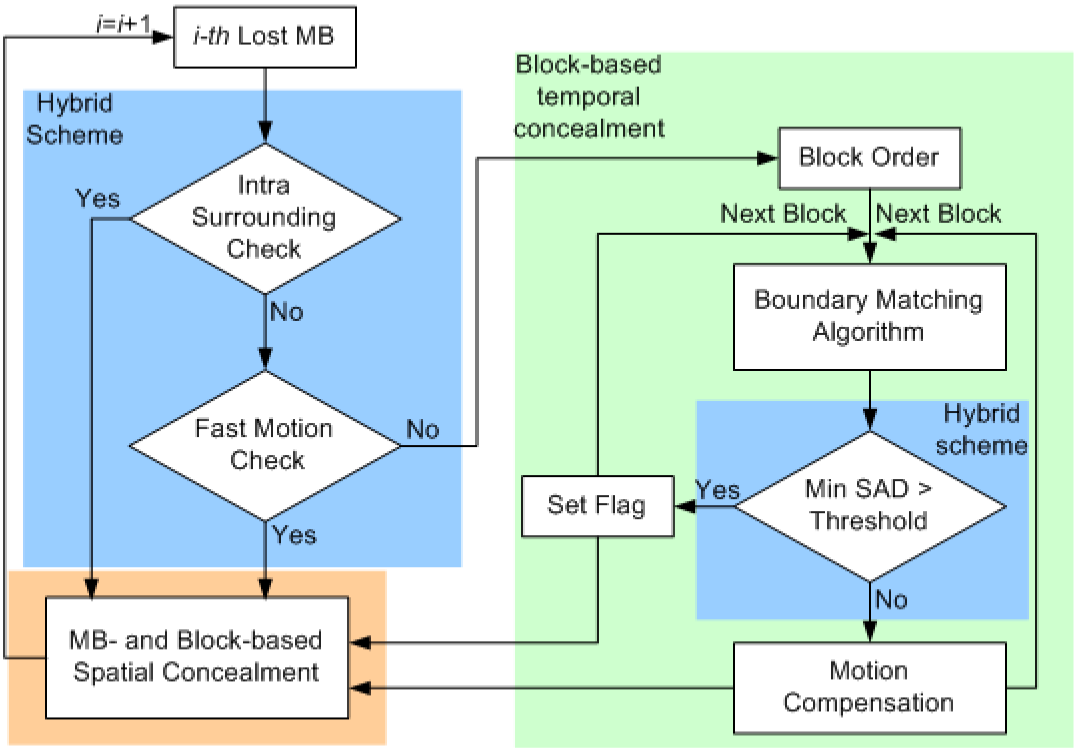


Fig. 29. Concealment process for lost MBs

3.2.2.3 Smoothing-Filter

Within a concealed frame, we may find some blocky effect caused by the different brightness of reconstructed MBs. To remove the blocky effect, a post filter is used. To investigate the visual quality improvement by the post filter, we use the deblocking filter of MPEG-4 Simple Profile reference software to reduce the blocky effect. The results are shown in Table 3. to Table 5. To observe the PSNR values of various frames, deblocked I-VOPs can have significant quality improvement in PSNR over the performance of non-processed I-VOPs. For P-VOPs, the deblocking filter usually decreases the averaged PSNR of the reconstructed video. By removal of the blocky effect, the processed P-VOPs can have better visual quality. In addition, for the low bitrate coding, the deblocking filter can get better performance for both P-VOPs and I-VOPs.

Table 3. PSNR for five sequences.

(I-frame/10frames)	Non-deblocking	deblocking
Foreman_256k	32.28	32.38
Akiyo_256k	41.5	41.49
Mobile_256k	24.29	24.33
Stefan_256k	25.19	25.38
Football_256k	28.74	29.1

Table 4. PSNR of Foreman sequence with constant QP.

(1^{st} I-VOP)	Non-deblocking	Deblocking
Foreman_QP4	38.04	37.52
Foreman_QP5	36.95	36.62
Foreman_QP9	33.6	33.59

Table 5. PSNR of Akiyo sequence with constant QP.

(1^{st} I-VOP)	Non-deblocking	Deblocking
Akiyo_QP2	44.89	44.52
Akiyo_QP3	43.15	43.11
Akiyo_QP4	41.2	41.25

To remove the blocking effect of the concealed VOPs, in addition to the deblocking filter in MPEG-4 reference software, we use two two-dimensional (2-D) filters of the first-order and second order 3x3 smooth filters as shown in Fig. 30.

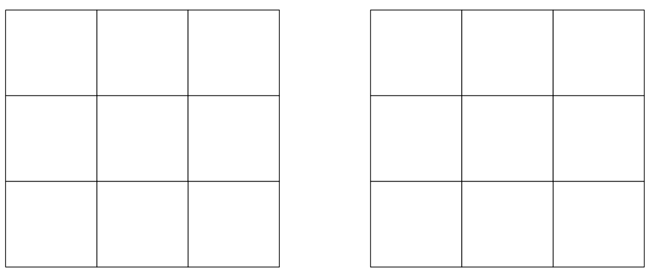


Fig. 30. First order and second order smooth filter

Because any inconsistency that occurs at the boundaries between the error MB and the neighbors are visible as the blocky artifact. In addition, the inconsistency is significant for low bitrate video coding. For video communications via channels with constrained bandwidth, the first-order 2-D filter can retain better visual quality than the second-order 2-D filter.

To have a reconstructed video of better visual quality, the authors in [4] rebuilt the lost MBs with the average values of neighboring three pixels instead of the interpolation values of the reference pixels, as in Fig. 31. The authors concealed the lost blocks with the average value of T1+T2+T3, or average value of D1+D2+D3, or average value of L1+L2+L3, or average value of R1+R2+R3, instead the average value of T2, D2, L2 and R2. The interpolation value for each pixel is computed by

$$P_1 = \frac{\frac{1}{3}(T1 + T2 + T3) * w2 + \frac{1}{3}(D1 + D2 + D3) * w1 + \frac{1}{3}(L1 + L2 + L3) * w4 + \frac{1}{3}(R1 + R2 + R3) * w3}{w1 + w2 + w3 + w4}.$$

Observing the above equation, we can combine the spatial concealment filter with the first-order 2-D filter to get identical reconstructed video with low complexity. The integration is defined by

$$P_2 = \frac{1}{9} * \left(\begin{array}{l} \left(\frac{T1 * (w2 + 1) + D1 * (w1 - 1) + L1 * (w4 + 1) + R1 * (w3 - 1)}{w1 + w2 + w3 + w4} \right) + \\ \left(\frac{T2 * (w2 + 1) + D2 * (w1 - 1) + L1 * w4 + R1 * w3}{w1 + w2 + w3 + w4} \right) + \\ \vdots \\ \left(\frac{T3 * (w2 - 1) + D3 * (w1 + 1) + L1 * (w4 - 1) + R1 * (w3 + 1)}{w1 + w2 + w3 + w4} \right) + \end{array} \right) = P_1$$

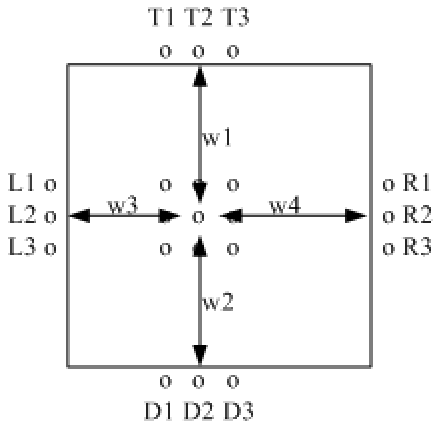


Fig. 31. The integration of the spatial interpolation and deblocking filtering.

3.3 Case Study: Hardware Acceleration

Due to the limited computation power and resources, parts of decoder workload need the coprocessors to share for throughput enhancement. According to the profile in Fig. 32, motion compensation (MC) and inverse discrete cosine transform (IDCT) take the significant computation time in decoding. Since IDCT has much regular data flow, a butterfly architecture in Fig. 33 is used to realize a fast IDCT. With 16 times of one-dimension (1-D) IDCT, including row and column operations, the IDCT of a MB can be finished in 18 cycles. The total hardware cost is 48,379.

The ARM platform contains RISC, logic module (FPGA), Static memory interface (SMI), ABMA bus and host interface. The designed co-processors is put in logic module and communicated through AHB (ARM high performance bus). ARM RISC runs 130 MHz and 10 MHz for logic module, 32bit 33 MHz for AHB. The experiment results are 8.1% throughput improvement in the decoder with hardware acceleration. Observing the simulation results, more speedups are required in the decoding rate to fit real-time requirements.

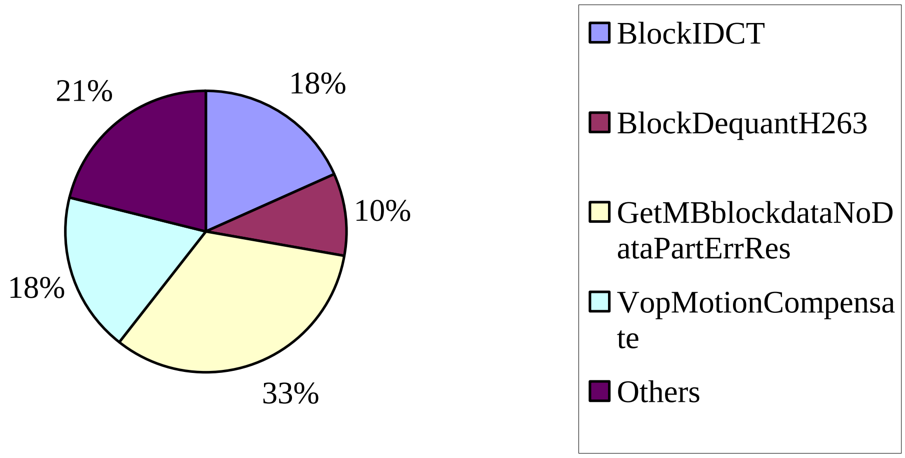


Fig. 32. The profiling of the NCTU Optimized Simple Profile MPEG-4 Decoder.

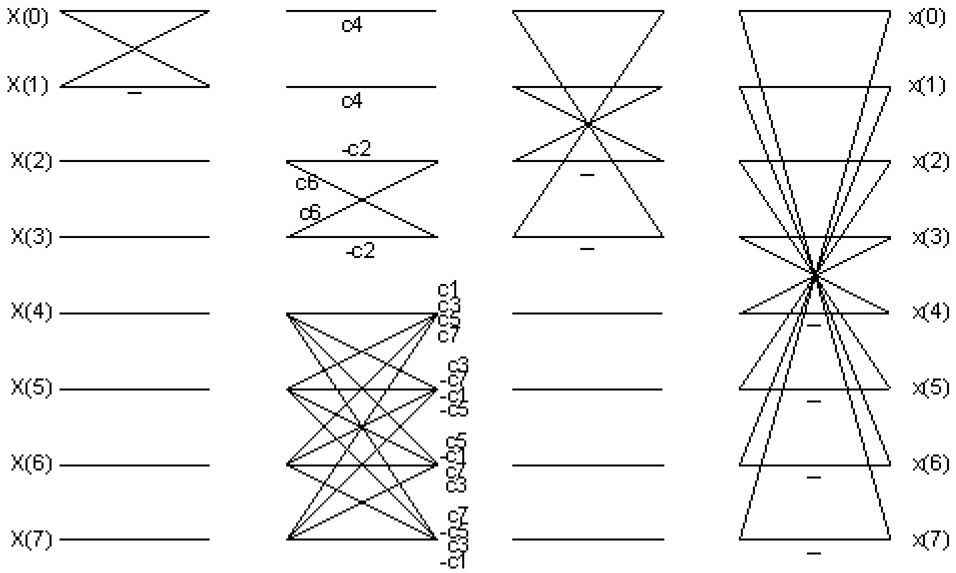


Fig. 33. The flow diagram of IDCT.

3.4 Summary

We propose an adaptive temporal/spatial error concealment to rebuild every lost block of the received video at the occurrence of the packet loss during transmission. An adaptive method to choose the best error concealment method by switching between spatial and temporal error concealment is proposed. In addition, a smooth filter is adopted to reduce the blocky effect, which can provide good visual quality. To deal with the VOP header loss, the time information is adopted for error detection and the direct copy approach is used to retain the temporal resolution of the received video. To fit real-time applications, the hardware accelerated IDCT module is developed. With the block-based error concealment, frame level error concealment, deblocking filtering and hardware acceleration, we can realize a real-time error resilience system for video streaming and communications.

Chapter 4

Experimental Results

To show the superiority of our proposed error concealment algorithms, both off-line testing and on-line testing are taken. In addition, for off-line testing, some encoding algorithms including intra refreshment are used to investigate the overall performance of an error resilience system. For on-line testing, retransmission scheme in the steaming system is taken into consideration for providing the better quality of service.

We use MPEG-4 Simple Profile reference software of MoMuSys version to implement our proposed error concealment algorithm.

4.1 Off-Line Testing

4.1.1 Experimental Environment

In order to test if our method can work on various sequence, we combined 20 different sequences into two long sequences to see the average performance. To see the performance of out concealment method in fast motion and slow motion, we take Foreman and Akiyo as our simulation sequences. To compare the performance of different resolution, we take a sequence of D1 resolution, which is most popular format for the digital video recorder now, to be simulated. The coding parameters are as follows: encoding frame rate is 30 frames/sec, decoding frame rate is 10 frames/sec, packet size is 2000 bits, GOP structure is I-P-P…, bitrate is 512k for normal test. To off-line simulate packet loss condition and see the effect of packet loss rate and concealment method, we use random drop with uniform distribution to simulate different packet loss rate. Because different lost places will make different results, we take the average of ten simulation results to get the average performance.

4.1.2 Error Concealment Benefit

First, we compare the benefit of each part in our method. We list the six tools which include different tool concealment schemes.

Table 6. Tool list

Tool 1 – simple	Spatial copy for I-VOP Zero motion for P-VOP
Tool 2 – BMA	Bi-directional Interpolation for I-VOP Temporal BMA for P-VOP
Tool 3 – BMA threshold	Hybrid scheme with BMA threshold based on tool 2
Tool 4 – fast motion	Fast motion checking based on tool 3
Tool 5 – surrounding intra	Surrounding intra checking based on tool 4
Tool 6 – block-based	Block-based refinement based on tool 5

Take Foreman for simulation, packet loss rate 10%, 700k bits/sec, the visual results are as Fig. 34.

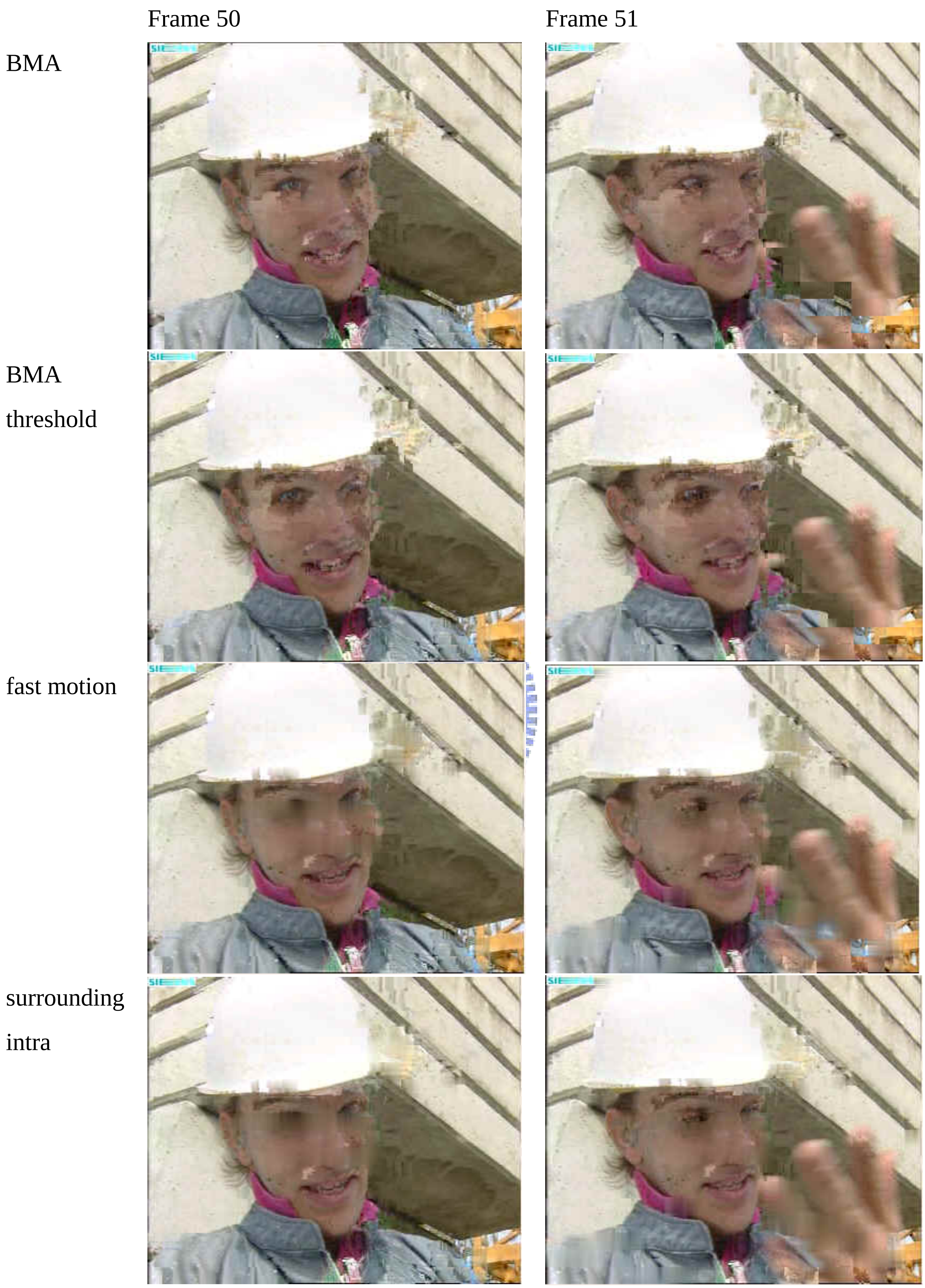


Fig. 34. Visual quality improvement of hybrid scheme benefit

To compare average performance of each proposed tool, different sequences are combined as following two tests. Sequence 1 combines 10 video sequences including Foreman, Akiyo, Bus, Football, Mobile, Stefan, Mother_daughter, Carphone, Coastquard, and Container. There are total 2400 frames.

Table 7. PSNR improvement of sequence 1

	Simple	BMA	BMA threshold	Fast motion	Surrounding intra	Block-based	Total Gain
PLR 1%	29.46	30.59	30.94	31.11	31.3	31.47	--
Gain	--	1.13	0.35	0.17	0.19	0.17	2.01
PLR 5%	22.01	24.07	24.84	24.96	25.24	25.61	--
Gain	--	2.06	0.77	0.12	0.28	0.37	3.6
PLR10%	19.53	21.87	22.5	22.77	23.12	23.52	--
Gain	--	2.34	0.63	0.27	0.35	0.4	3.99

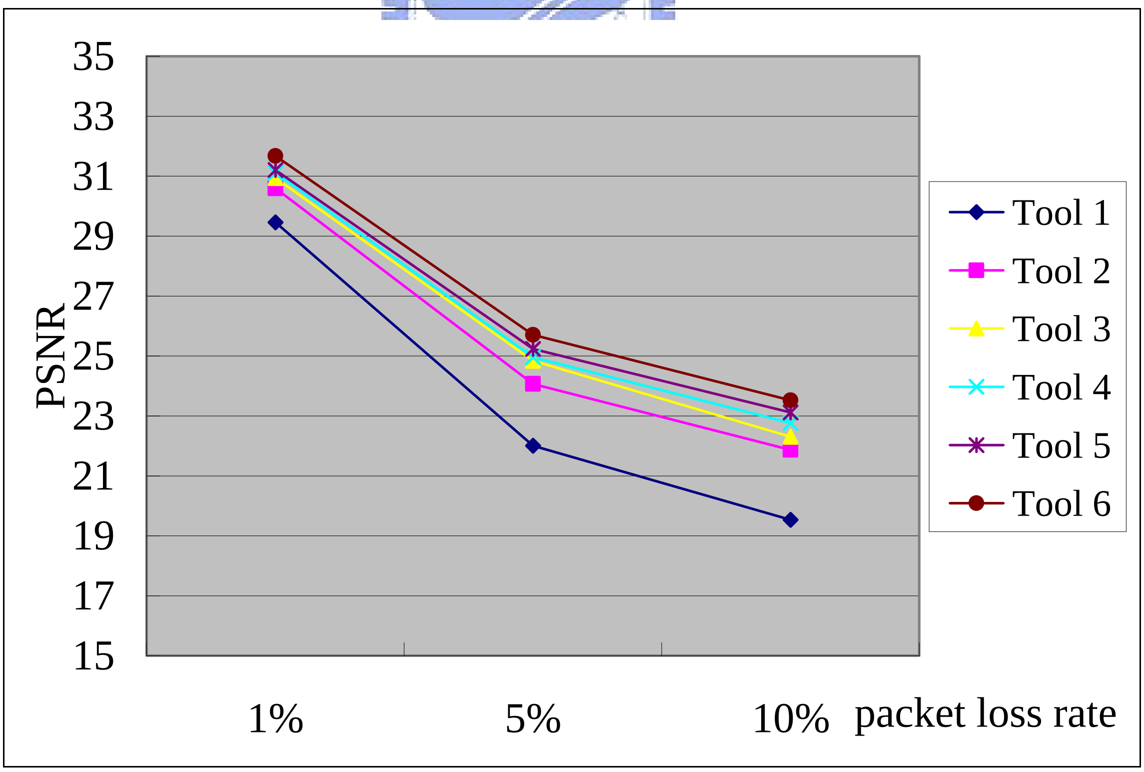


Fig. 35 PSNR curve of sequence 2 by different tools on different packet loss rate

Sequence 2 combines 10 video sequences including Dancer, Funfair, Hall_monitor, News, Paris, Silent, Singer, Table, Tempete, and Waterfall. There are total 2900 frames.

Table 8. PSNR improvement of sequence 2

	Simple	BMA	BMA threshold	Fast motion	Surrounding intra	Block-based	Total Gain
PLR 1%	29.52	30.4	30.78	30.92	31.08	31.31	--
Gain	--	0.88	0.38	0.14	0.16	0.23	1.79
PLR 5%	22.02	23.93	24.38	24.56	24.81	25	--
Gain	--	1.91	0.45	0.18	0.25	0.19	2.98
PLR10%	18.73	20.81	22.41	22.66	22.83	23.19	--
Gain	--	2.08	1.6	0.25	0.17	0.36	4.46

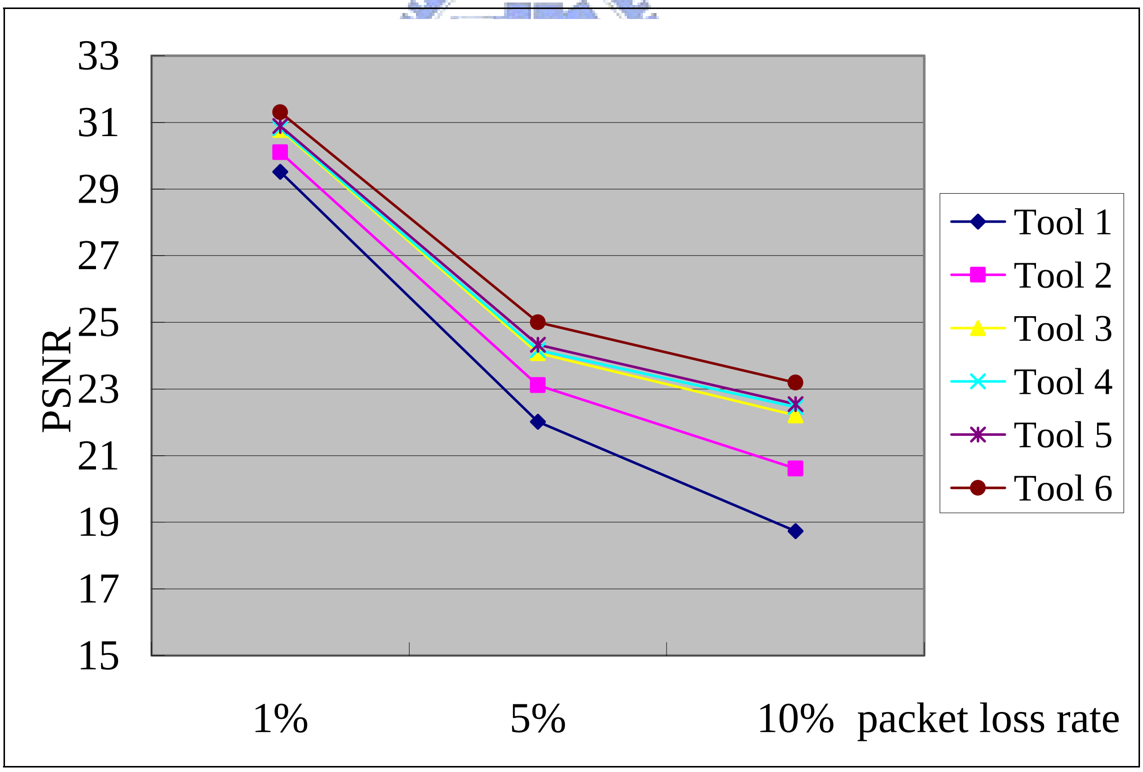


Fig. 36. PSNR curve of sequence 2 by different tools on different packet loss rate

Fig. 37 shows the performance the smoothing filter. The Foreman sequence is encoded at 700k bits/sec. Packet loss rate is 10%. Frame 33 of the reconstructed video is illustrated.

Without smooth With smooth

Fig. 37. Visual quality of using smoothing filter

4.1.3 Motion and Resolution Tests

As for slow motion test, we take Akiyo to be simulated at different bitrate and packet loss rate. We can get 0.86~6.32 dB gain.

Table 9. PSNR improvement of slow motion simulation

Akiyo									
bitrate(bits/sec)	50k			100k			200k		
	Tool 6	Tool 1	Gain	Tool 6	Tool 1	Gain	Tool 6	Tool 1	Gain
Error Free	37.63	--	--	40.34	--	--	42.76	--	--
PLR 1%	36.05	35.19	0.86	37.22	36.25	0.97	38.96	35.93	3.03
PLR 5%	31.75	27.11	4.64	33.3	29.77	3.53	32.6	26.28	6.32
PLR 10%	28.61	23.15	5.46	29.62	24.32	5.3	30.87	24.65	6.22

As for fast motion test, we take Foreman to be simulated at different bitrate and packet loss rate. We can get 0.39~5.39 dB gain.

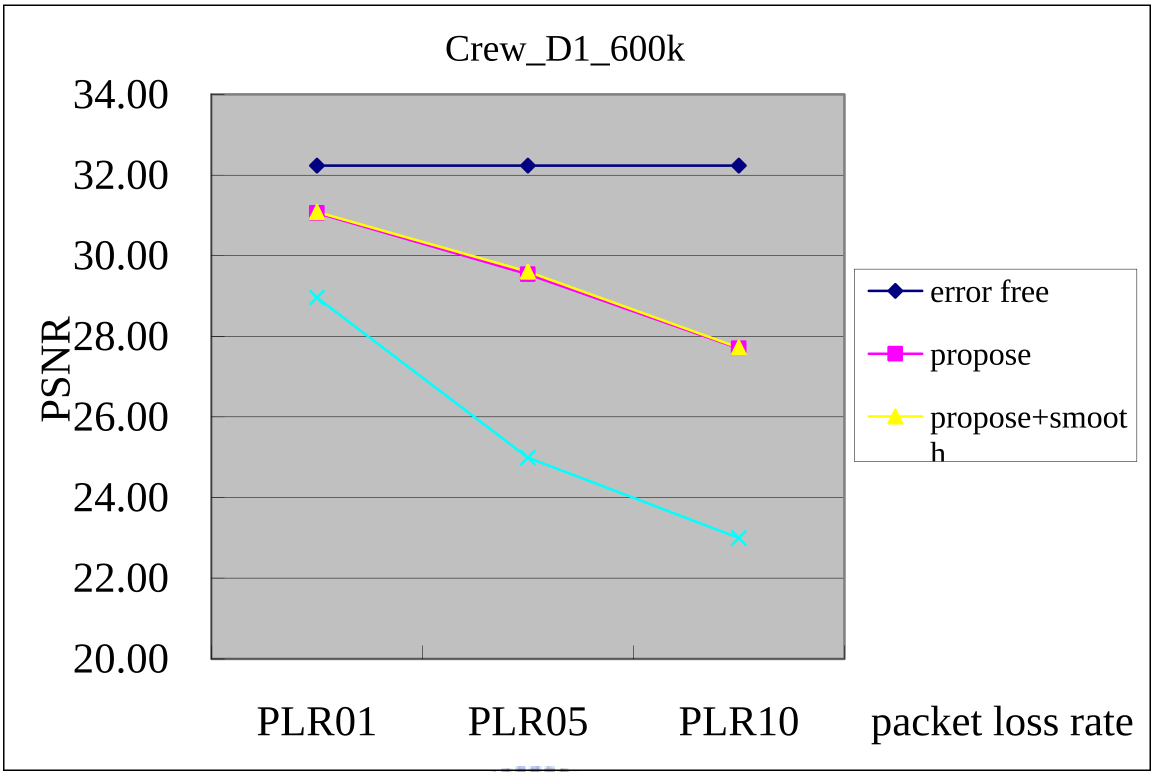
Table 10. PSNR improvement of fast motion simulation

Foreman									
bitrate(bits/sec)	128k			256k			512k		
	Tool 6	Tool 1	Gain	Tool 6	Tool 1	Gain	Tool 6	Tool 1	Gain
Error Free	30.28	--	--	33.65	--	--	36.9	--	--
PLR 1%	28.01	27.62	0.39	31.14	28.73	2.41	32.3	27.05	5.25
PLR 5%	25.31	22.66	2.65	26.59	21.88	4.71	27.53	22.14	5.39
PLR 10%	23.04	19.31	3.73	23.78	19.02	4.76	24.91	19.54	5.37

As for D1 resolution test, we take Crew to be simulated at different bitrate and packet loss rate. We can get 0.52~3.62 dB gain.

Table 11. PSNR improvement of D1 resolution simulation

Crew									
bitrate(bits/sec)	600k			900k			1200k		
	Tool 6	Tool 1	Gain	Tool 6	Tool 1	Gain	Tool 6	Tool 1	Gain
Error Free	27.08	–	–	29.13	–	–	30.86	–	–
PLR 1%	26.74	26.22	0.52	28.51	27.76	0.75	30	29.08	0.92
PLR 5%	25.45	23.65	1.8	26.66	23.71	2.95	27.54	24.85	2.69
PLR 10%	24.25	21.47	2.78	25.07	21.04	4.03	25.62	22	3.62



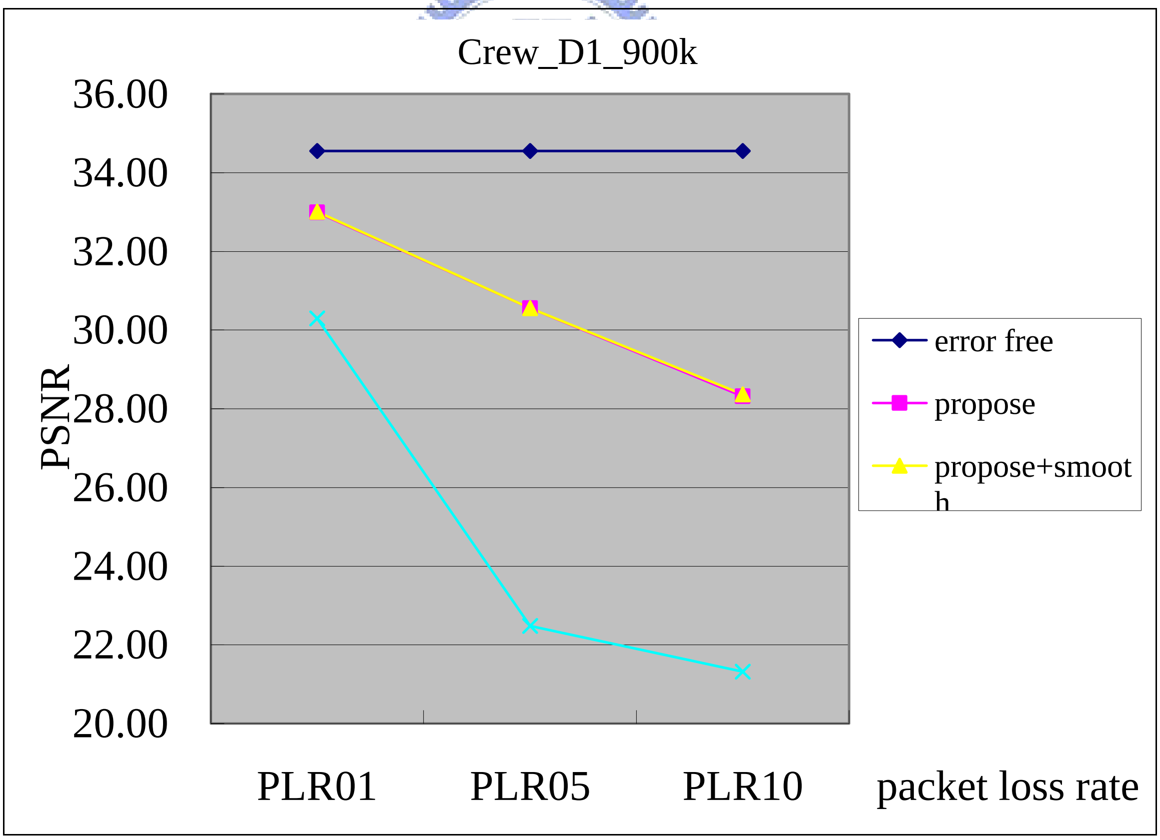


Fig. 38. Quality comparisons of the different error concealment schemes at packet loss rates including 1% (PLR01), 5% (PLR05) and 10% (PLR10).

4.1.4 Combination with Intra Refreshment

To evaluate the performance of the error resilience system, 3 series of experiments are adopted. For VOD applications under various network conditions, the error resilient capabilities are examined at 3 bitrates including 256, 550 and 700 kilo-bits per sec. In each experiment, 4 error resilience systems as listed in Table 12 are compared. Type 1 represents the original reference encoder and decoder system with default error resilience tools including Resync_Markers and zero-motion-vector spatial copy. Type 2 enables intra refresh algorithm [23] in the encoder. Type 3 enables the proposed error robustness and error concealment algorithm. Type 4 is the error system with intra refreshment in the encoder and our proposed error concealment in the decoder. 3 different network conditions including packet loss rate (PLR) of 1%, 5% and 10% with uniform dropping probability model are used for testing. Test sequence is encoded with one I-VOP's and 99 P-VOP's with the sequences used.

The test results are shown in Table 13 to Table 18. We can find that the Type 2 can get a gain over Type 1 with 3.64-8.45 dB in PSNR. Type 3 can achieve a gain over Type 1 with 5.16-10.62 dB in PSNR. Based on error resilient decoder and rate-distortion intra refresh encoder, we can get a gain over Type 1 with 6.33-13.22 dB in PSNR. Fig. 39 to Fig. 44 demonstrates rate-distortion (R-D) curves to compare performance of the four error resilience systems with the different packet loss rates (PLR).

Fig. 45 shows the quality improvement over the other 3 types and the objective quality is much better.

Table 12. The 4 types of system with embedded ER tools.

Type	Encoder	Decoder
1	Resynchronization marker	Zero motion for P-VOPs and spatial copy for I-VOPs
2	Intra-refresh	Zero motion for P-VOPs and spatial copy for I-VOPs
3	Resynchronization marker	Proposed hybrid concealment
4	Intra-refresh	Proposed hybrid concealment

Table 13. The reconstructed image quality for Foreman with 260 kbps.

		Type 1	Type 2	Type 3	Type 4
PLR[2]	PSNR[3]	33.59	-	-	-
	Gain[1]	-	-	-	-
1%	PSNR	28.26	30.22	30.92	31.70
	Gain[1]	0	1.96	2.66	3.44
5%	PSNR	21.80	26.69	26.60	28.65
	Gain[1]	0	4.89	4.8	6.85
10%	PSNR	19.03	23.67	23.78	25.78
	Gain[1]	0	4.64	4.75	6.75

[1]Gain: the difference compared to PSNR of type 1 (unit:dB)

[2]PLR : packet loss rate

[3]unit: dB

Table 14. The reconstructed image quality for Foreman with 550 kbps.

PLR2		Type 1	Type 2	Type 3	Type 4
0%	PSNR3	36.94	-	-	-
	Gain1	-	-	-	-
1%	PSNR	28.87	32.78	31.90	34.55
	Gain1	0	3.91	3.03	5.68
5%	PSNR	22.00	28.00	26.94	30.75
	Gain1	0	6.00	4.94	8.75
10%	PSNR	19.64	26.32	24.55	28.75
	Gain1	0	6.68	4.91	9.23

Table 15. The reconstructed image quality for Foreman with 700 kbps.

PLR2		Type 1	Type 2	Type 3	Type 4
0%	PSNR3	38.04	-	-	-
	Gain1	-	-	-	-
1%	PSNR	30.10	34.29	34.03	35.70
	Gain1	0	4.19	3.93	5.6
5%	PSNR	21.60	28.46	27.05	31.63
	Gain1	0	6.86	5.45	10.03
10%	PSNR	19.41	26.74	25.23	29.54
	Gain1	0	7.33	5.82	10.13

Table 16. The reconstructed image quality for Akiyo with 130 kbps.

PLR[2]		Type 1	Type 2	Type 3	Type 4
0%	PSNR[3]	41.2	-	-	-
	Gain[1]	-	-	-	-
1%	PSNR	22.57	39.83	33.86	39.95
	Gain[1]	0	17.26	11.29	17.38
5%	PSNR	21.68	37.5	30.18	39.21
	Gain[1]	0	15.82	8.5	17.53
10%	PSNR	20.62	35.55	28.03	36.6
	Gain[1]	0	14.93	7.41	15.98

Table 17. The reconstructed image quality for Akiyo with 200 kbps.

PLR[2]		Type 1	Type 2	Type 3	Type 4
0%	PSNR[3]	43.15	-	-	-
	Gain[1]	-	-	-	-
1%	PSNR	22.61	41.34	33.93	42.1
	Gain[1]	0	18.73	11.32	19.49
5%	PSNR	21.7	38.68	30.12	40.13
	Gain[1]	0	16.98	8.42	18.43
10%	PSNR	20.63	36.28	28.78	37.38
	Gain[1]	0	15.65	8.15	16.75

Table 18. The reconstructed image quality for Akiyo with 310 kbps.

PLR[2]		Type 1	Type 2	Type 3	Type 4
0%	PSNR[3]	44.89	-	-	-
	Gain[1]	-	-	-	-
1%	PSNR	22.55	39.83	34.48	43.74
	Gain[1]	0	17.28	11.93	21.19
5%	PSNR	21.72	37.5	31.02	40.59
	Gain[1]	0	15.78	9.3	18.87
10%	PSNR	20.64	35.55	29.26	37.75
	Gain[1]	0	14.91	8.62	17.11

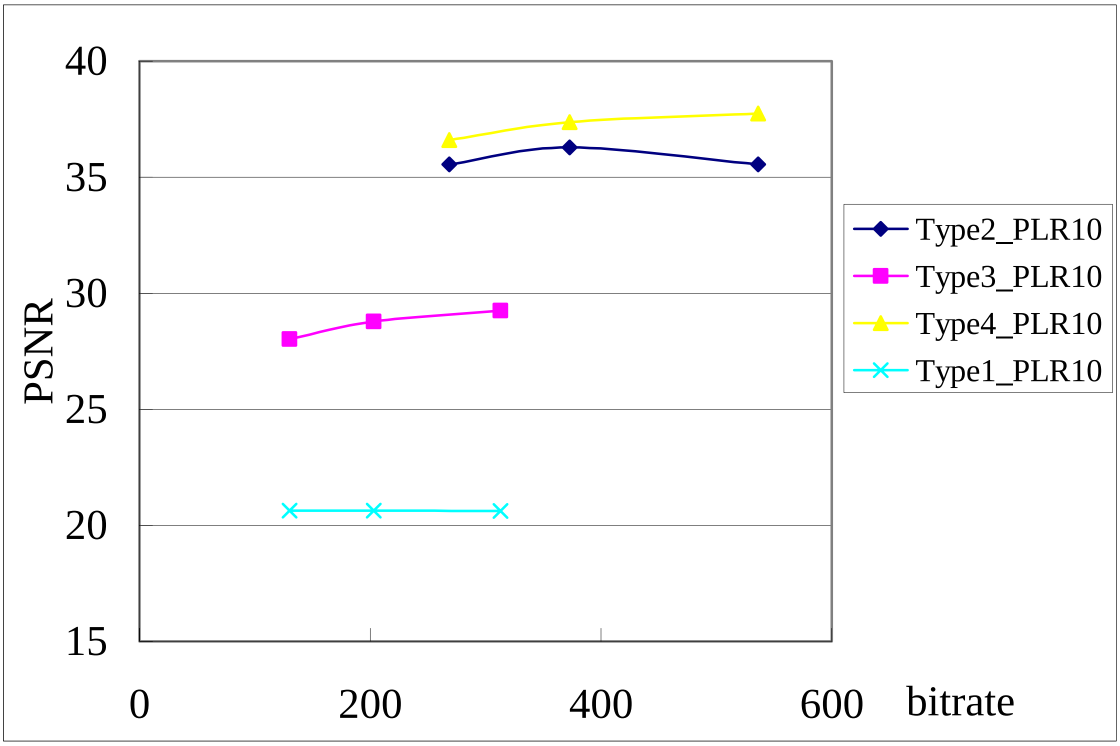


Fig. 39. R-D curve of reconstructed image quality for Akiyo with PLR 10%.

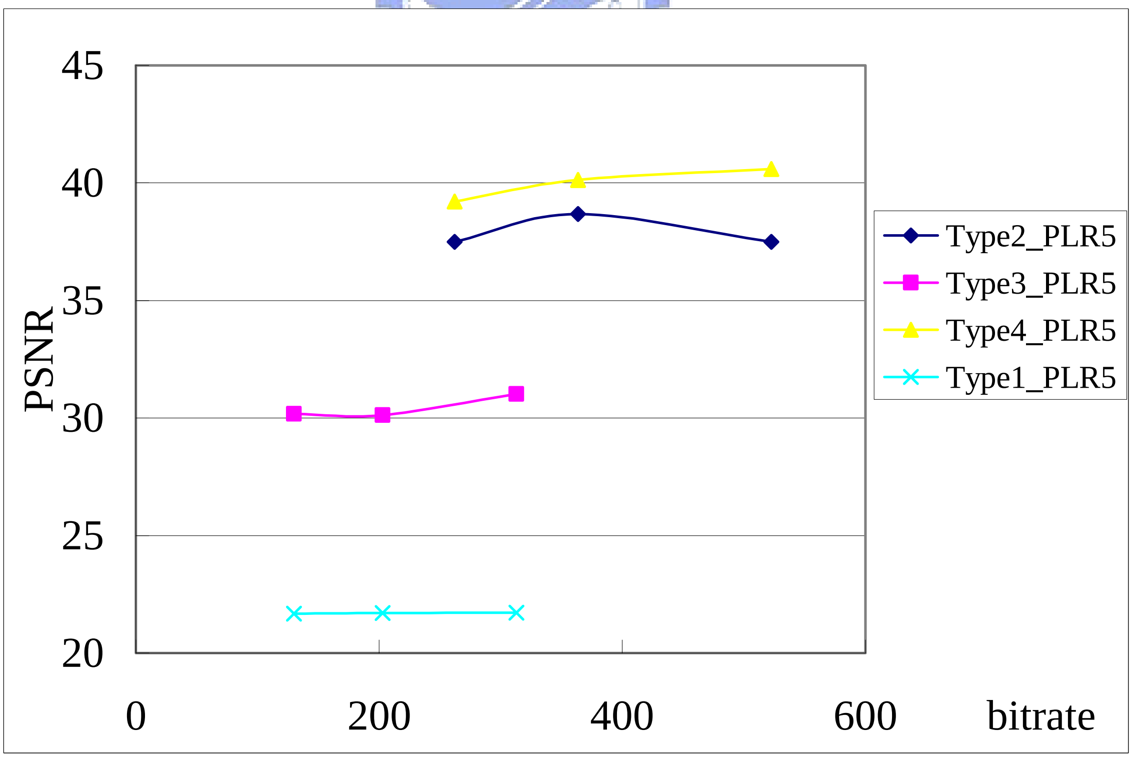


Fig. 40. R-D curve of reconstructed image quality for Akiyo with PLR 5%.

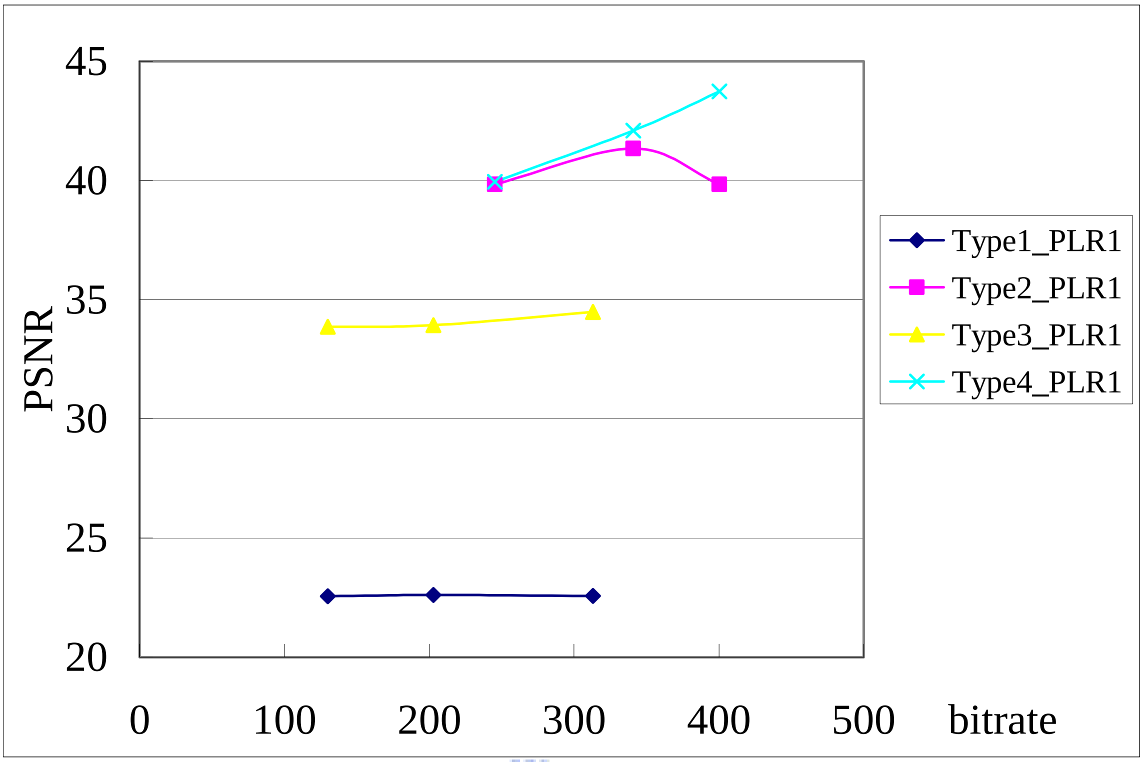


Fig. 41. R-D curve of reconstructed image quality for Akiyo with PLR 1%.

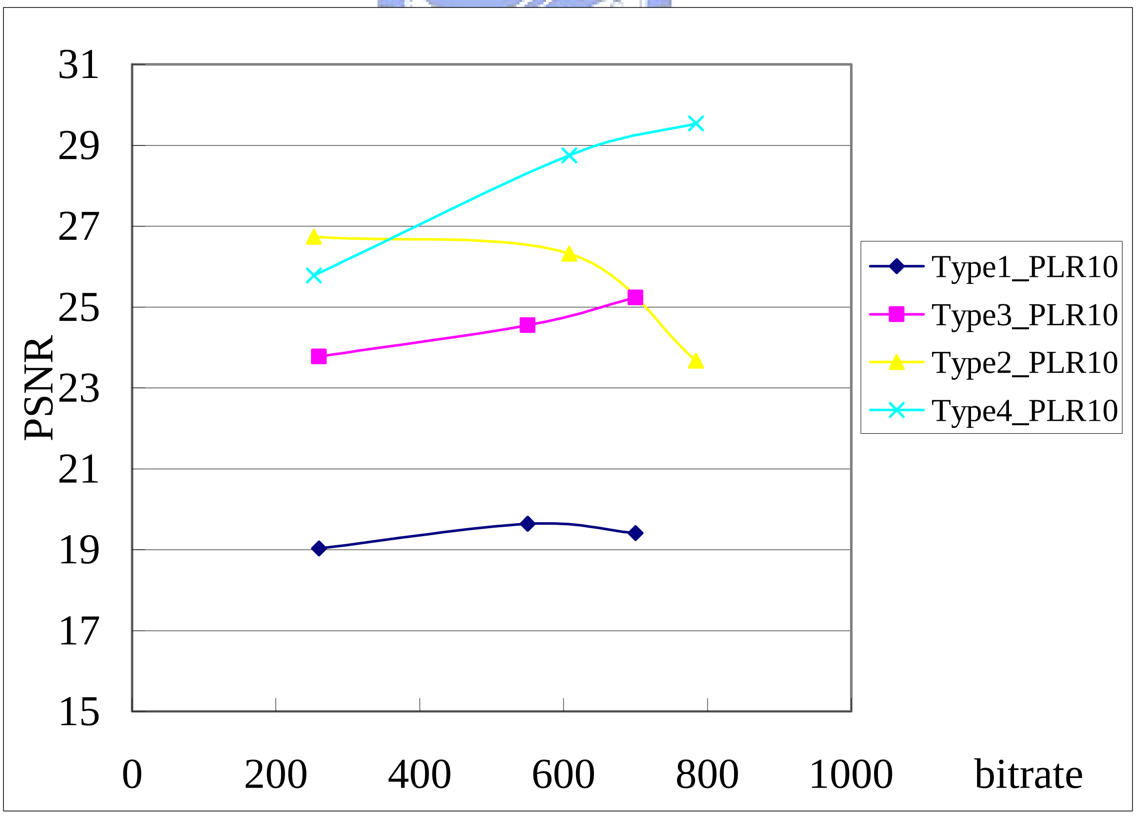


Fig. 42. R-D curve of reconstructed image quality for Foreman with PLR 10%

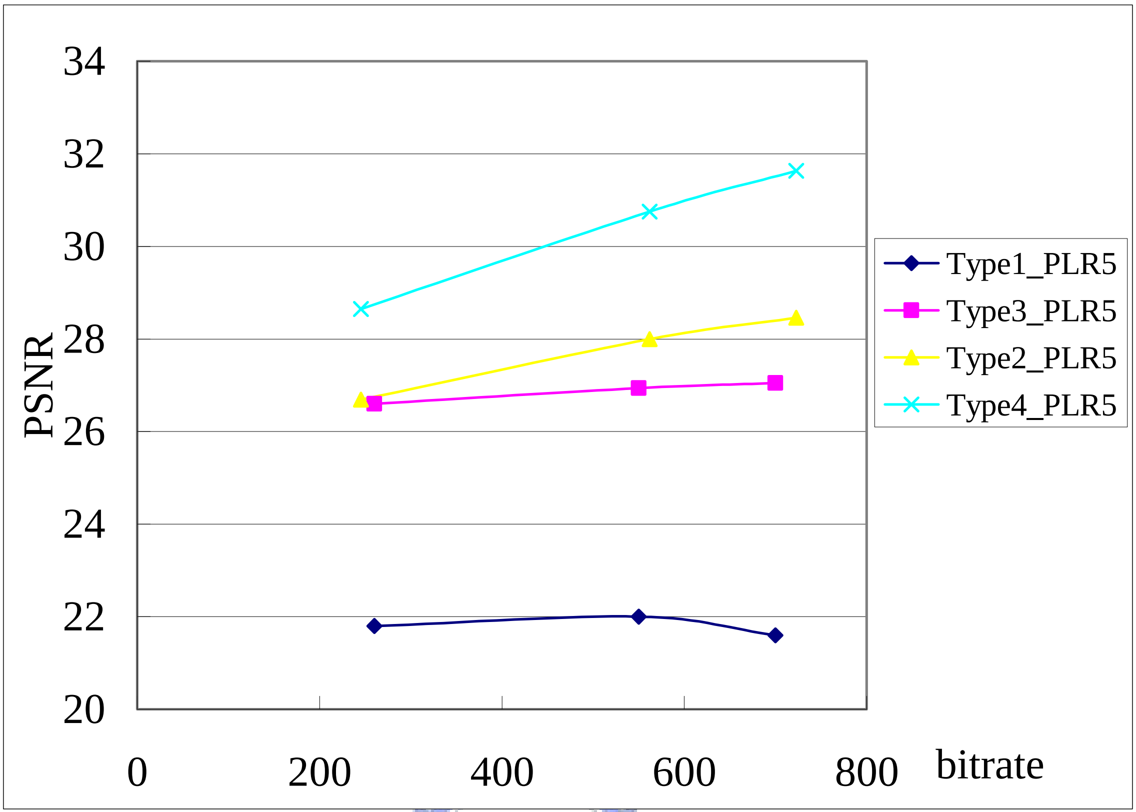


Fig. 43. R-D curve of reconstructed image quality for Foreman with PLR 5%

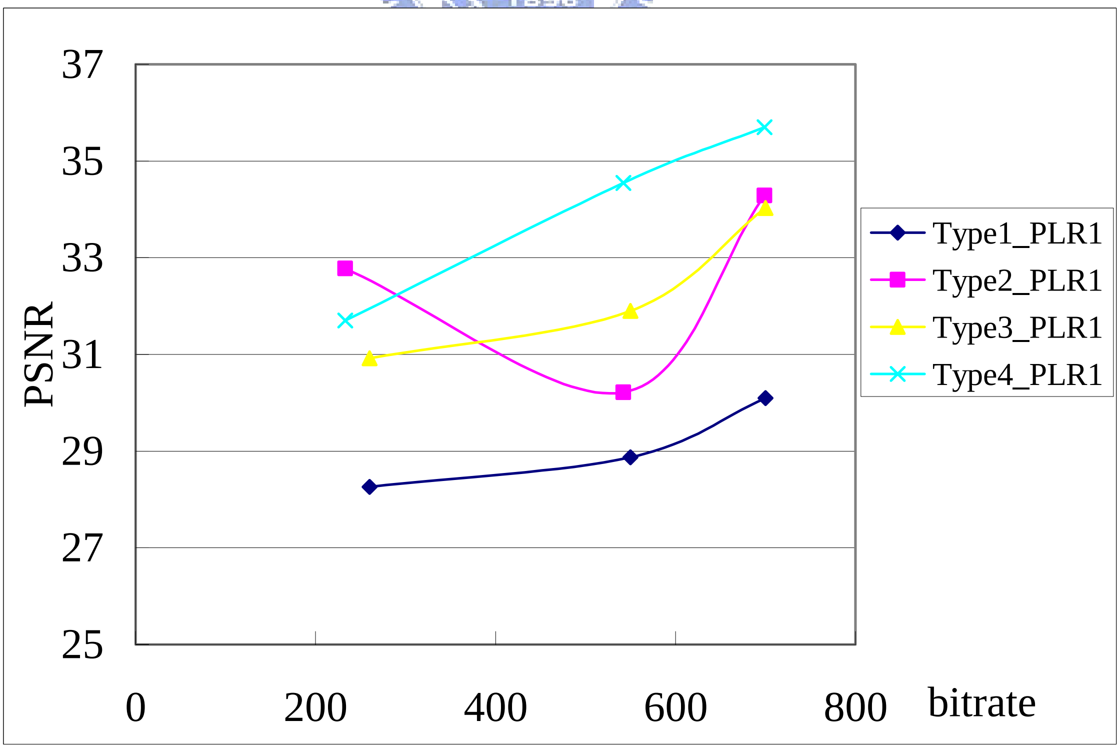


Fig. 44. R-D curve of reconstructed image quality for Foreman with PLR 1%

Fig. 39 to Fig. 44 show a tendency that the PSNR values at high bitrate are lower than the PSNR values at low bitrate under the use of identical error resilient tools and the same packet loss rate. For off-line testing, we use random model to drop the bistream packets. The concealment performance is related with the lost area. When dropping the detailed or important areas, we may get the worse concealed video as any packet loss.

Fig. 45. The image quality comparison for the 4 types of error resilience system. (a) Type 1; (b) Type 2 ;(c) Type 3 ;(d) Type 4. (No. 8th frame, PLR 10%, bitrate 700kbps)

4.2 On-Line Testing

With the on-line evaluation on the multimedia test bed, the simulations of video content delivery with the proposed robust system without use of motion information over wireless channels are shown in Fig. 46 and Fig. 47. We test each case with different combinations of a retransmission scheme and Resync_Markers to evaluate the visual quality of the reconstructed video sequences. Tool 1 disables simultaneously the retransmission scheme of the test bed and the use of Resync_Marker in the coded bitstream. Tool 2 adopts the same techniques as in the Tool 1 except the use of Resync_Markers. Tool 3 enables both tools at the same time, which can have the best subjective quality on the average for streaming video contents over the wireless channels. The simulation results show that the Tool 3 has the best visual quality. As observed from the reconstructed video frames, the proposed video streaming system can provide better subjective visual quality of receiving videos for the consumers.

(a)

(a)

(b)

(c)

Fig. 46. The visual quality comparisons for the 3 combinations of the tools within the error resilience system. (a) Tool 1; (b) Tool 2; (c) Tool 3.

(b)

(c)

Fig. 47. The visual quality comparisons for the 3 combinations of the tools within the error resilience system (a) Tool 1; (b) Tool 2; (c) Tool 3.

As the packet loss rate increase, retransmission can reduce the packet loss rate of transmission channel. Deciding which packet should be retransmitted first or using efficient retransmission can make the results better. Besides, network bandwidth is also an impact factor affecting the retransmission performance. With enough bandwidth to retransmit all the packets, we could make the decoder to get the error free bitstream.

4.3 Summary

The experiment results show that the proposed error concealment method and error resilience system can improve the video quality under the packet loss through transmission channels. In addition, intra refreshment in the encoder for off-line testing or retransmission scheme in the system for on-line testing is also taken into consideration. By combining the motion concealment, intra refreshment and retransmission, the error resilient and crash proof decoder can improve the visual quality of received video significantly.

Chapter 5

Conclusion

5.1 Contributions

This thesis has developed an efficient error concealment and error resilient MPEG-4 decoding system.

1. For any packet loss that causes the loss of MBs or VOP headers during the transmission, we have proposed an error resilient decoding system that can prevent the decoder from being crashed, provide a reconstructed video of proper visual quality, and retain the temporal resolution.
2. For the corrupted MBs, we have proposed an adaptive block-based hybrid error concealment method. The innovations are the hybrid concealment scheme with less complexity, block-based concealment to refine MB-based method, and a smooth filter is used to make the concealment results have better visual quality. Based on our proposed method, the damaged video can recover the quality and provide a good performance both on PSNR and visual quality.
3. The error resilient decoding system is simulated on the MPEG-21 part-12 multimedia test bed under various network conditions. The simulation results show that the error resilient decoding system is applicable to video communications or video streaming.

5.2 Future Works

For the multimedia including text, audio, and video, error resilience only on video coding is not sufficient to provide a good media service. Different kinds of media all need to be investigated. For the streaming system, decoder is combined with client. Taking client and network conditions into consideration, we may have more information and ability to get best performance for multimedia communication.

References

[1] J. Zhang, J. F. Arnold and M. R. Frater, "A cell-loss concealment technique for MPEG-2 coded video", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 10, No. 4, pp.659-665, June 2000.

[2] B. Yan, and K. W. Ng, "A Novel Selective Motion Vector Matching Algorithm for Error Concealment in MPEG-4 Video Transmission over Error-Prone Channels", *IEEE Transactions on Consumer Electronics*, Vol. 49, No. 4, pp.1416-1423, November 2003

[3] L. Atzori, F. G. B. De Natale, and C. Perra, "A Spatio-Temporal Concealment Technique Using Boundary Matching Algorithm and Mesh-Based Warping (BMA–MBW)", *IEEE Transactions On Multimedia*, Vol. 3, No. 3, pp.326-338, September 2001

[4] S. Valente, C. Dufour, F. Groliere, and D. Snook, "An Efficient Error Concealment Implementation for MPEG-4 Video Streams", *IEEE Transactions on Consumer Electronics*, Vol.47, No.3, pp.568-578, August 2001

[5] M. Ghanbari and V. Seferidis, "Cell-Loss Concealment in ATM Video", *IEEE transactions on circuits and systems for video technology*, Vol.3, No.3, pp.238-247, June 1993

[6] S. Kaiser, K. Fazel, "Comparison of error concealment techniques for an MPEG-2 video decoder in terrestrial TV-broadcasting", *Signal Processing: Image Communication*, Vol. 14, No.6-8, pp. 655-676, May 1999

[7] S. Cen and P. C. Cosman, "Decision Trees for Error Concealment in Video Decoding", *IEEE Transactions On Multimedia*, Vol. 5, No. 1, pp.1-7, March 2003

[8] W.-J. Chu and J.-J. Leou, "Detection and Concealment of Transmission Errors in H.261 Images", *IEEE Transactions On Circuits And Systems For Video Technology*, Vol. 8, No. 1, pp.74-84, February 1998

[9] M.-H. Jo and W.-J. Song, "Error Concealment for MPEG-2 Video Decoders with Enhanced Coding Mode Estimation", *IEEE Transactions on Consumer Electronics*, Vol.46, No.4, pp.962-969, November 2000

[10] P. Salama, N. B. Shroff, and E. J. Delp, "Error Concealment in MPEG Video Streams Over ATM Networks", *IEEE Journal On Selected Areas In Communications*, Vol. 18, No. 6, pp.1129-1144, June 2000

[11] J.-W. Suh and Y.-S. Ho, "Error Concealment Techniques for Digital TV", *IEEE Transactions On Broadcasting*, Vol. 48, No. 4, pp.299-306, December 2002

[12] L. D.-Soares, F. Pereira, "Error resilience and concealment performance for MPEG-4 frame-based video coding", *Signal Processing: Image Communication*, Vol. 14, No.6-8, pp.447-472, May 1999

[13] J. F. Arnold, M. R. Frater, J. Zhang, "Error resilience in the MPEG-2 video coding standard for cell based networks - A review", *Signal Processing: Image Communication*, Vol.14, No.6-8, pp.607-633, May 1999

[14] Y. Wang, S. Wenger, J. Wen, and A. K. Katsaggelos, "Error Resilient Video Coding Techniques ", *IEEE Signal Processing Magazine*, Vol.17, No.4, pp.61-82, July 2000

[15] M. E. A.-Mualla, C.N. Canagarajah and D. R. Bull, "Motion Field Interpolation for Temporal Error Concealment", *IEE Processing Vision, Image and Signal Processing*, Vol.147, No.5, pp.445-453, October 2000

[16] S. Tsekeridou, and I. Pitas, "MPEG-2 Error Concealment Based on Block-Matching Principles", *IEEE Transactions On Circuits And Systems For Video Technology*, Vol. 10, No. 4, pp.646-658, June 2000

[17] W. Kwok and H. Sun, "Multi-Directional Interpolation for Spatial Error Concealment", *IEEE Transactions on Consumer Electronics*, Vol.39, No.3, pp.455-460, August 1993

[18] J.-Y. Pyun, J.-S. Lee, J.-W. Jeong, J.-H. Jeong, and S.-J. Ko, "Robust Error Concealment for Visual Communications in Burst-Packet-Loss Networks", *IEEE Transactions on Consumer Electronics*, Vol. 49, No. 4, pp.1013-1019, November 2003

[19] S. Aign, K. Fazel, "Temporal and Spatial Error Concealment Techniques for Hierarchical MPEG-2 Video Codec", *1995 IEEE International Conference on Communications*, Vol.3, pp.1778-1783, 18-22 June 1995

[20] E. T. Kim, S.-J. Choi, H.-M. Kim, "Weighted boundary matching algorithm for error concealment in the MPEG-2 video bit stream", *Signal Processing*, Vol.73, No.3, pp.291-295, January 3,1999

[21] Y.-L. Lee, C.-N. Wang and T. Chiang, "An MPEG-4 error resilient decoder", *2002 Taiwan Workshop on Consume Electronics*, October 2002

[22] M.-Y. Huang, T.-L. Su, S.-H. Wang, C.-N. Wang, and T. Chiang, "An Error Resilient System for Streaming MPEG-4 Video over the Internet", *2003 Taiwan Workshop on Consume Electronics*, November 27-28, 2003

[23] T.-L. Su, M.-Y. Huang, H.-C. Chuang, C.-N. Wang, and T. Chiang, "A Robust MPEG-21 Media Delivery Video Streaming System", *8th World Multiconference on Systemics, Cybernetics and Informatics (SCI2004)*, July 18-21, 2004

[24] H. Sun, J. W. Zdepski, W. Kwok, and D. Raychaudhuri, "Error Concealment Algorithm for Robust Decoding of MPEG Compressed Video", *Signal Processing：Image Communication*, Vol.10, No.4, pp.249-268, September 1997

[25] M.-C. Hong, H. Schwab, L. P. Kondi, and A. K. Katsaggelos, "Error concealment algorithms for compressed video," *Signal Processing: Image Communication*, vol. 14, no. 6-8, pp. 473 -492, May 1999

[26] R. Talluri, I. Moccagatta, Y. Nag, and G. Cheung, "Error concealment by data partitioning", *Signal Processing: Image Communication*, Vol.14, No.6-8, pp.505-518, May 1999

[27] M. Bystrom, V. Parthasarathy, and J. W. Modestino, "Hybrid Error Concealment Schemes for Broadcast Video Transmission over ATM Networks", *IEEE Transactions On Circuits And Systems For Video Technology*, Vol. 9, No. 6, pp.868-881, September 1999

[28] I. Burnett, R. V. de Walle, K. Hill, J. Bormans, and F. Pereira, "MPEG-21: Goals and Achievements", *IEEE Multimedia*, Vol.10, No.4, pp.60-70, October–December 2003

[29] J. Bormans, J. Gelissen, and A. Perkis, "MPEG-21：The 21st Century Multimedia Framework", *IEEE Signal Processing Magazine*, Vol.20, No.2, pp. 53-62, March 2003

[30] J. R. Smith, "MPEG-21 Digital Item Adaptation : Enabling Universal Multimedia Access", *IEEE Multimedia*, Vol.11, No.1, pp.84-87, January–March 2004

[31] A. Perkis, J. Zhang, T. Halvorsen, J. O. Kjode, and F. Rivas, "A Streaming Media Engine Using Digital Item Adaptation", *2002 IEEE Workshop on Multimedia Signal Processing*, pp.73-76, 9-11 December 2002

[32] C.-N. Wang, C.-Y. Tsai, H.-C. Lin, H.-C. Chuang, Y.-C. Lin, J.-H. Chen, K. L. Tong, F.-C. Chang, C.-J. Tsai, S.-Y. Lee, T. Chiang, and H.-M. Hang, "FGS-Based Video Streaming Test Bed for MPEG-21 Universal Multimedia Access with Digital Item Adaptation", *in Proc. of ISCAS2001*, 2001.

自　　　傳

黃名彥: 民國 1980 年生於高雄市。 2002 畢業於台灣新竹的國立交通大學電子工程學系，之後進入該校電子工程所攻讀碩士學位。以 MPEG-4 解碼器為論文研究主題。

Ming-Yen Huang was born in KaoHsiung in 1980. He received the BS degree in Electronics Engineering, National Chiao Tung University (NCTU), HsinChu, Taiwan in 2002. His current research interest is MPEG-4 decoder.