ABSTRACT

Title of Dissertation: ERROR RESILIENCE IN
HETEROGENEOUS VISUAL COMMUNICATIONS

Meng Chen, Doctor of Philosophy, 2007

Dissertation directed by: Professor Min Wu
Department of Electrical and Computer Engineering

A critical and challenging aspect of visual communication technologies is to immunize visual information to transmission errors. In order to effectively protect visual content against transmission errors, various kinds of heterogeneities involved in multimedia delivery need to be considered, such as compressed stream characteristics heterogeneity, channel condition heterogeneity, multi-user and multi-hop heterogeneity. The main theme of this dissertation is to explore these heterogeneities involved in error-resilient visual communications to deliver different visual content over heterogeneous networks with good visual quality.

Concurrently transmitting multiple video streams in error-prone environment faces many challenges, such as video content characteristics are heterogeneous, transmission bandwidth is limited, and the user device capabilities vary. These challenges

prompt the need for an integrated approach of error protection and resource allocation. One motivation of this dissertation is to develop such an integrated approach for an emerging application of multi-stream video aggregation, i.e. multi-point video conferencing. We propose a distributed multi-point video conferencing system that employs packet division multiplexing access (PDMA)-based error protection and resource allocation, and explore the multi-hop awareness to deliver good and fair visual quality of video streams to end users.

When the transport layer mechanism, such as forward error correction (FEC), cannot provide sufficient error protection on the payload stream, the unrecovered transmission errors may lead to visual distortions at the decoder. In order to mitigate the visual distortions caused by the unrecovered errors, concealment techniques can be applied at the decoder to provide an approximation of the original content. Due to image characteristics heterogeneity, different concealment approaches are necessary to accommodate different nature of the lost image content. We address this heterogeneity issue and propose to apply a classification framework that adaptively selects the suitable error concealment technique for each damaged image area.

The analysis and extensive experimental results in this dissertation demonstrate that the proposed integrated approach of FEC and resource allocation as well as the new classification-based error concealment approach can significantly outperform conventional error-resilient approaches.

ERROR RESILIENCE IN HETEROGENEOUS VISUAL COMMUNICATIONS

by

Meng Chen

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2007

Advisory Committee:

Professor Min Wu, Chair/Advisor
Professor Rama Chellappa
Professor Zhi-Long Chen
Professor K. J. Ray Liu
Professor André L. Tits

DEDICATION

To Yi and Jonathan.

ACKNOWLEDGEMENTS

I am extremely grateful to my advisor, Min Wu, for being a great mentor for my Ph.D study. Her critical insights guided me through my course of research. Her patience and encouragement motivated me to overcome hardships in the journey to become a researcher. Completing a Ph.D program is a great mile-stone for my life and she made it happen.

I am thankful for having the opportunity to work with a couple of distinguished researchers, Dr. Guan-Ming Su and Dr. Yefeng Zheng. Working with them has greatly benefited my research work and shaped my viewpoint of how to conduct valuable research. I would like to thank Prof. André L. Tits, Prof. Zhi-Long Chen, Prof. K. J. Ray Liu, and Prof. Rama Chellappa, for being my thesis committee members and giving insightful suggestions to my research work.

I also would like to thank all the group mates and friends during my study at University of Maryland at College Park, for their time and effort to directly work with me or provide suggestions to my research work.

Finally, I would like to express my earnest gratitude to my husband, Yi, and my son, Jonathan. I could not accomplish my research work without Yi's strong support. And Jonathan, my little angel, makes everyday of my life joyful.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ARQ	Automatic Repeat Request
DCT	Discrete Cosine Transform
FEC	Forward Error Correction
FGS	Fine Granularity Scalability
GOP	Groups of Pictures
GSB	Geometric-Structure-Based
I frame	Intra Frame
JPEG	Joint Photographic Experts Group
KKT condition	Karush-Kuhn-Tucker condition
LRC	Long Range Correlation
MCU	Multi-Point Control Unit
MD-FEC	Multiple Descriptions through Forward Error Correction
MDI	Multi-Directional Interpolation
MPEG	Moving Picture Experts Group
MSE	Mean Square Error
MSR	Maximally Smooth Recovery
OASI	Orientation Adaptive Sequential Interpolation
P frame	Forward Predicted Frame
PDMA	Packet Division Multiplexing Access
PLR	Packet Loss Rate
POCS	Projection-Onto-Convex-Sets
PSNR	Peak Signal to Noise Ratio
QoS	Quality of Service
R-D	Rate-Distortion

RS code	Reed-Solomon code
RSVP	Resource Reservation Protocol
S-D	Segment-Distortion
SVM	Support Vector Machine
TDMA	Time Division Multiplexing Access
UEP	Unequal Error Protection

Chapter 1

Introduction

1.1 Motivation

In recent years, we have witnessed a phenomenal growth of digital visual communications. Multimedia bit-stream can be damaged during transmission because of channel error conditions or bandwidth limitation. Due to the temporal and spatial prediction in video/image compression, erroneously received samples of compressed bit-stream can cause the distortion of large portions of visual content. To immunize visual information to transmission errors becomes a critical and challenging aspect of visual communication technologies. The demand for such technologies has been accelerated by a large amount of multimedia service deployments over various types of networks.

There are two main classes of error-resilient techniques at transport level: error detection plus retransmission and Forward Error Correction (FEC). Error detection allows a receiver to check whether the received data has been corrupted during transmission, so that a request for a retransmission could be initiated if needed. FEC allows a receiver to reconstruct the original information by introducing information redundancy. In this dissertation, we focus on error-resilient systems employing FEC

because FEC can be applied in applications with real-time constraints, where error detection plus retransmission is less suitable. When the transport layer mechanism, such as FEC, cannot provide sufficient error protection on the payload stream, the unrecovered transmission errors may lead to visual distortions at the decoder. To mitigate the visual distortions caused by the unrecovered errors, concealment techniques can be applied at the decoder to provide an approximation of the original content.

Both FEC and error concealment have received a lot of attention in the research community in recent years [67, 59, 52, 76, 35] and have become widely deployed. However, emerging multimedia service deployment scenarios involving multi-stream video aggregation pose new challenges for error-resilient techniques. One major challenge for effectively protecting visual content against transmission errors is adapting to various kinds of heterogeneities that affect the performance of error resilience, including compressed stream characteristics, channel condition, multi-user, and multi-hop heterogeneities. It requires a major research effort to model and investigate the effective error-resilient techniques and efficient resource allocation strategies to explore these multiple dimensions of heterogeneities. The main theme of this dissertation is to analyze, model, and solve error resilience problems in heterogeneous visual communications to deliver different visual content over heterogeneous networks with good visual quality. As an example, various types of highly demanded multimedia services, such as video conferencing, video gaming and remote teaching, may involve concurrently transmitting multiple video streams in an error-prone environment, where video content characteristics are heterogeneous, the user device capabilities vary, and networks are heterogeneous. These challenges prompt the need for an integrated FEC and resource allocation approach which

explores multi-stream, multi-user and network heterogeneities. One motivation of this dissertation is to develop such an integrated approach for multi-stream video aggregation in emerging applications of multi-point video conferencing. We propose a distributed multi-point video conferencing system and a packet division multiplexing access (PDMA)-based error protection scheme which is employed in distributed devices/nodes to minimize the maximal expected video distortion among all aggregated streams. In order to achieve good and fair visual quality of all delivered video streams to all end users in our distributed multi-point video conferencing system, we further explore the multi-hop awareness for multi-stream video aggregation over packet erasure channels.

In general, the transport level mechanisms may not provide sufficient protection for all the visual content. The unrecovered transmission errors may lead to visual distortions at the receiver. In this case, error concealment could be performed at the receiver to reconstruct the loss information. The widely used block-based visual coding systems have prompted a need of block-based error concealment on the decoder side. If contiguous image blocks are assembled in the same packet, the loss of one packet results in the loss of contiguous image blocks. It makes the recovery of the lost image blocks more difficult. One strategy to overcome this defect is block interleaving [74]. With block interleaving, the loss of a packet only affects noncontiguous image blocks. The spatial concealment approach using surrounding pixels information of a lost block is then an effective technique to reconstruct the damaged visual content. A number of such concealment approaches have been proposed in recent years [67, 76, 35, 69, 60, 68, 78, 3]. The smoothness and continuity properties in spatial or frequency domain, the repeating patterns, and other properties of visual data have been exploited to recover corrupted blocks from the survived

surroundings. We have observed that different approaches are suitable for different image characteristics of a corrupted block and its surroundings, and none of the existing approaches is an all-time champion. This motivates us to explore the image characteristics heterogeneity for error concealment technique. We found that the concealment quality could be substantially improved if we could intelligently combine state-of-the-art approaches. The classification technique acts of taking in the raw data and making a decision on the "category" of pattern [19]. It naturally fits as a technique to achieve our objective. In this dissertation, we propose using classification to integrate state-of-the-art error concealment techniques and adaptively select the suitable algorithm for each damaged image area.

In summary, for a communication system to be effectively resilient to transmission errors, error-resilient techniques should be investigated by exploring various types of heterogeneities involved in visual content transmission. Our research effort involves knowledge in several scientific areas such as resource allocation, optimization, and classification, and demonstrates promising frameworks for multimedia error-resilient approaches. The analysis and extensive experimental results in this dissertation show that our proposed FEC and resource allocation integrated approach, and classification-based error concealment approach can significantly outperform conventional error-resilient approaches by exploring content characteristics, channel condition, multi-user, and multi-hop heterogeneities.

1.2 Dissertation Organization

The dissertation presents an integrated framework on FEC and resource allocation for multi-stream video aggregation in Chapters 2–5 and a classification-based error concealment framework to accommodate image characteristics heterogeneity

in Chapter 6.

We propose a distributed framework for realizing multi-point video conferencing and a packet division multiplexing access (PDMA)-based error protection scheme in Chapter 2. PDMA-based error protection scheme is then modeled as an optimization problem to minimize the maximal expected video distortion among all aggregated streams in Chapter 3. In Chapter 4, we propose an algorithm to reach preference consensus for all conferees to accommodate the user preference heterogeneity in a multi-point video conferencing system. In order to achieve good and fair visual quality of all delivered video streams to all end users in our distributed multi-point video conferencing system, we further explore the multi-hop awareness in Chapter 5 for multi-stream video aggregation over packet erasure channels and propose an optimal error protection and resource allocation algorithm. In Chapter 6, we look into how the image characteristics heterogeneity affects the performance of different error concealment approaches, and address this issue by proposing classification-based error concealment framework. Finally, we conclude this dissertation and discuss some future perspectives in Chapter 7.

Chapter 2

Distributed Conferencing System

Transmitting real-time encoded video streams over various types of networks has been enabled by the rapid development of video coding, communications, and multimedia display technologies. One emerging application is multi-point video conferencing, which realizes a virtual conference room for multiple participants located at different geographical areas. There are several design challenges for a multi-point video conferencing system over packet erasure channels, where the video quality could be severely degraded due to packet loss. First, as each conferee transmits his/her real-time compressed video stream through resource-limited links, proper resource allocation among multiple video streams is important. Second, the channel conditions in different hops and characteristics of different video streams are inherently heterogeneous. The optimal error protection for different streams along various hops may not be the same. An effective error protection solution should be able to adapt to multi-stream multi-hop heterogeneity, and apply error protection accordingly. Furthermore, a multi-point video conferencing requires real-time streaming, whereby a strict delay constraint is imposed to each stream to maintain the interactivity within a conference session. This demands a solution with short de-

lay for parameters exchange. In this chapter, we address the aforementioned issues and propose a distributed multi-point video conferencing system.

A simple realization of multi-point video conferencing can be implemented by each user sending multiple unicasting streams to all other conferees. In addition to the inefficiency caused by transmitting redundant copies of video content, it is difficult for each user to perform timely error protection for each stream to achieve the optimal video quality, subject to the time-varying channel condition within each channel, heterogeneous conditions along the streaming path, and a long feedback delay for end-to-end channel condition.

Although multicasting can alleviate redundant copies of multiple streams, realizing a multi-point video conferencing via multicasting still requires obtaining timely channel information for end-to-end channel condition and needs to consider the heterogeneity of channel conditions experienced by all video streams. Optimization approaches have been proposed for resource allocation in a multicast session. They can be performed either on the sender side [44, 30] or in a receiver-driven manner [47, 64, 66]. However, these approaches may not be able to provide good and fair visual quality for all video conferencing attendees. This is due to the unawareness of existence of other users' streams that are aggregated through the same communication link in different multicast sessions. The communication system is likely to reserve the same bandwidth for each stream aggregated over a communication link in different multicast sessions. Given varying content complexity in different video streams, using the same bit-rate for all streams could result in undesirably low quality for some video streams and unnecessarily high quality for other video streams which are displayed in low-resolution[62]. This motivates us to develop a multi-point video conferencing system, which explores multi-stream heterogeneity

to address the error protection and resource allocation challenges.

Conventional centralized multi-point video conferencing system could be considered to explore the multi-stream heterogeneity in terms of bandwidth allocation. It often deploys a centralized scheme controlled by a multi-point control unit (MCU). In general, centralized multi-point video conferencing system may encounter the problem of resource bottleneck at MCU when the number of conferees increases or the complexity of processing algorithm for each video stream increases. Therefore, it often focuses on error-free communication channel [63, 17, 38, 20] without considering error protection for each video stream. In addition, the centralized system cannot react to the fast changing conditions in both channel and video source owing to long delay in information exchange. Supporting information exchange locally and performing joint source and channel coding in a distributed manner can speed up the reaction to varying channel conditions and adapt to the heterogeneous conditions in different hops [55, 54]. By doing so, we arrive at a distributed design for a multi-point video conferencing system.

Compared to the transmission of generic data and voice in a distributed system [8], providing real-time video conferencing service in a distributed system is more involving. For example, the commonly used variable bit-rate compression for video poses more difficulties on the network resource allocation than voice transmission where constant bit-rate compression is generally used. Furthermore, the compressed video bit-streams exhibit decoding dependency on the previous coded bit-streams owing to the spatial and temporal prediction. The part of video stream where corrupted bits cause severe error propagation should have stronger error protection applied than the rest of video stream. In this chapter, we propose a multi-point video conferencing system by aggregating multiple streams with unequal error protection

in a distributed manner. Unlike the traditional time division multiplexing access (TDMA)-based error protection approach that allocates several complete packets for each video stream's source and parity check symbols, we propose a packet division multiplexing access (PDMA)-based allocation by allowing each packet to carry all video streams' source symbols and the parity check symbols. The analytical studies show that PDMA-based error protection has superior performance comparing to TDMA-based approach.

This chapter is organized as follows. We introduce a distributed multi-point video conferencing system in Section 2.1. In Section 2.2, we describe the building blocks of the proposed conferencing system. Error protection schemes for multi-stream video aggregation are proposed in Sections 2.3. The analytical studies of comparing TDMA-based and PDMA-based error protection schemes are presented in Section 2.4 and chapter summary is presented in Section 2.5.

2.1 System Overview

In this section, we present an architectural overview of a distributed multi-point video conferencing system. This distributed multi-point video conferencing system for S participating users is illustrated in Fig. 2.1. There are two types of nodes in this system, namely, user node (UN) and video combiner node (CN). Both types of nodes contain two components, namely, a video source module and a resource allocation module. To transmit a video stream from one user node to all other user nodes, there are three different kinds of links involved, namely, user node to video combiner ($UN_s - CN$) through channel U_s, video combiner to video combiner ($CN_m - CN_n$) through channel C_{mn}, and video combiner to user node ($CN - UN_s$) through channel V_s. Here, s is the user node index, and m and n

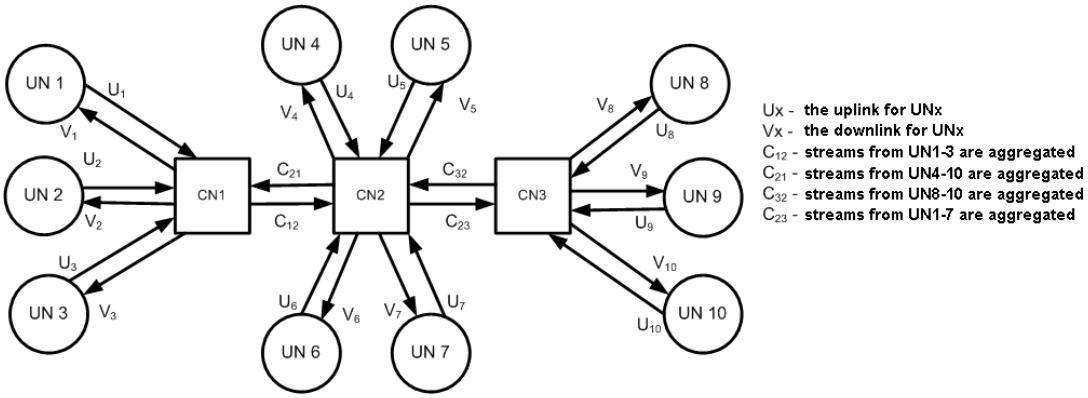


Figure 2.1: Proposed system topology of a distributed multi-point video conferencing system. UN stands for user node and CN stands for video combiner node.

are video combiner node indices. Without loss of generality, we consider the case where one video combiner node serves as a "portal" for a user node. Since the channel condition is time varying and feedback through multiple hops may introduce undesirable delay for real-time applications, it is often difficult for each CN node to be aware of communication links' channel condition in a conferencing system other than its transmission channel condition. Therefore, in our distributed video conferencing system, each CN node performs multi-stream aggregation locally by applying resource allocation and error protection based on its transmission channel condition and the rate-distortion (R-D) information of its aggregated video streams.

Specifically, in $(UN_s - CN)$ transmission, the video source module located in UN_s captures the video and analyzes the R-D information of the video content for each incoming video frame. The channel information of outbound link U_s is obtained through feedback channel from CN to UN_s . The resource allocation module performs joint optimization by selecting the parameters for source coding and channel coding based on the R-D and channel information. The compressed video streams

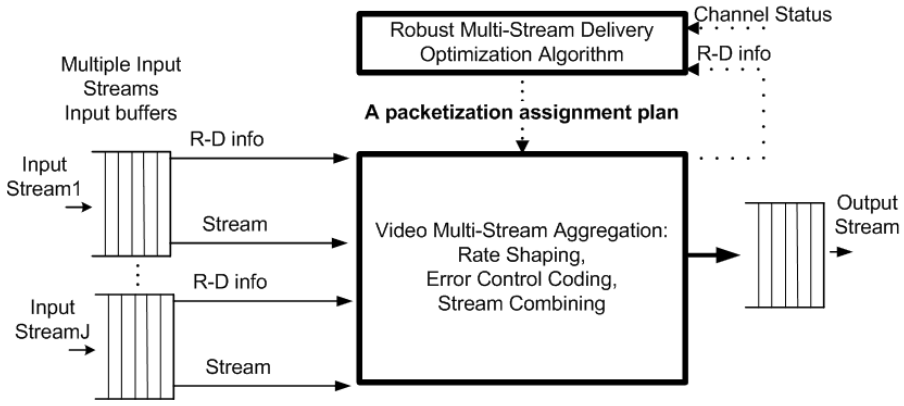


Figure 2.2: The multi-stream aggregation scheme for the resource allocation module in a video combiner node.

embedded with R-D information are then transmitted to the corresponding video combiner.

For $(CN_m - CN_n)$ transmission, the video source module located in the transmitter node CN_m buffers all incoming coded video streams from different users through UN-CN transmission. After collecting one-frame video data, the video source module performs channel decoding to obtain the video source bit-streams and the corresponding R-D information for each video stream. In addition, the channel information of link C_{mn} is obtained from the feedback of next-hop video combiner CN_n. Then, the resource allocation module located at video combiner CN_m performs multi-stream optimization to jointly select the parameters of source coding and channel coding for all incoming streams, and transmits this protected and merged stream to the next video combiner. $(CN - UN_s)$ transmission is similar to $(CN_m - CN_n)$ transmission, except that the receiver is a user node instead of a video combiner node. User s receives all the other users' video bit-streams from the nearest video combiner through link V_s.

In all three types of transmissions, the resource allocation module basically performs the same joint multi-stream operation, namely, based on video source R-D of all incoming streams and transmission channel information to perform multi-stream resource allocation optimization. Fig. 2.2 shows a diagram of the proposed multi-stream aggregation scheme for the resource allocation module in video combiner CN_m. The scheme first buffers the incoming video source bit-streams and obtains the R-D information from video source module. The resource allocation module performs a joint source/channel optimization and determines a packetization assignment plan for the incoming video streams. After these video streams are packetized according to the plan, they are moved to the output buffer for transmission. The $(UN_s - CN)$ transmission can be treated as a special case with only one incoming stream.

2.2 Building Blocks: Source and Channel Coding

The video codec for our system should provide high flexibility to facilitate rate adaptation and provide accurate R-D information with low overhead. We adopt MPEG-4 Fine Granularity Scalability (FGS) coding [53, 41, 46] in this work to demonstrate the concept, while the proposed framework can be extended to incorporate other scalable codecs. MPEG-4 FGS is a two-layer scheme consisting of a non-scalable base-layer and a highly scalable FGS enhancement layer. Its enhancement layer for each frame can be truncated at any point to achieve the desired rate, and the corresponding video quality decreases gracefully with the reduction in rate. We refer to this enhancement layer as the FGS-layer in this dissertation. The R-D function of FGS-layer at the frame level can be well approximated as a piecewise linear line by interpolating the R-D pairs obtained for recovering each complete DCT

bitplane [77]. Therefore, the R-D function for each video frame can be described using a small amount of bits.

For error control, we use forward error correction (FEC) codes because it can be applied to applications with real-time constraints, for which the approach of error detection plus retransmission is less suitable. A widely deployed FEC code is Reed-Solomon (RS) code, which achieves the upper bound for the minimum distance of an (n, k) linear code with code-word length n and the code-word dimension k. In addition to this Singleton bound [39], "shortening" is another fine property of the RS code to produce a code-word of any desired size by deleting some symbols from a RS code-word. The minimum distance of the shortened code-words still achieves the Singleton bound. This property provides RS code the capability of easily adapting to desired packet size for packet-based communication protocols. As video conferencing applications commonly deploy packet-based communication protocols and errors are primarily due to packet loss, we adopt the RS code for error recovery.

We focus on the fixed-length packetization because it is a relatively matured technique and widely used for its simplicity. The FEC coding and packetization for a single video stream can be achieved as follows: Let L be the number of symbols carried in a packet and N be the total number of packets. A segment is defined as the set of symbols located at the same position of each of the N packets. For the non-scalable base-layer, a strong, equal error protection strategy is applied to ensure its delivery as shown in Fig. 2.3(a). Because FGS enhancement layer uses bitplane based coding [53], the decoding of the symbols in its remaining bitplanes following a lost symbol may not improve the visual quality of the received video bit-stream. Therefore, FGS enhancement data has a monotonically decreasing priority

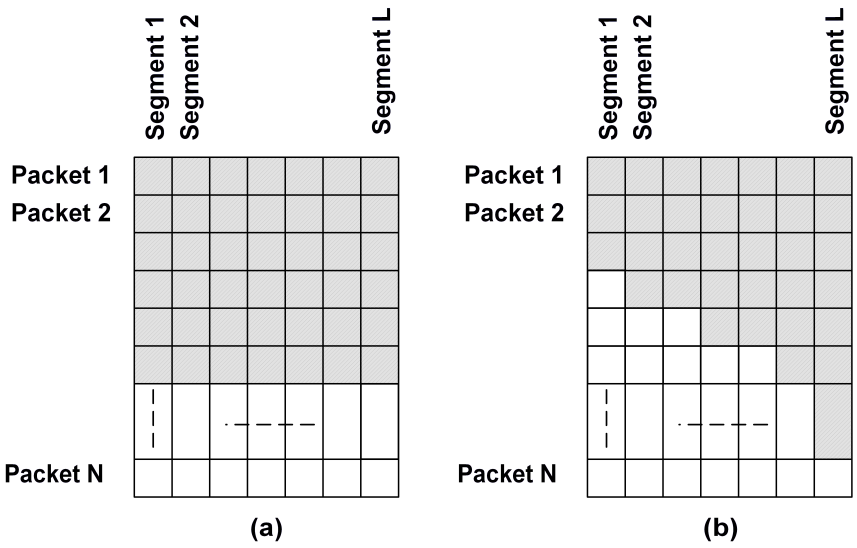


Figure 2.3: Single video stream protection by RS codes. White part indicates the RS code symbols. The shaded part indicates the source symbols. RS code is applied segment (column) by segment (column). (a) Equal error protection. (b) Unequal error protection.

for error protection. We consider an unequal error protection method of multiple descriptions through forward error correction codes (MD-FEC) [52] for FGS-layer, which has been shown to achieve good perceptual video quality in delivering single video stream. Given N packets, MD-FEC fills the FGS bit-stream vertically into N packets segment by segment in a stair case fashion as shown in Fig. 2.3(b), and Reed-Solomon (RS) code is applied within each segment. A higher error protection level of RS code is applied for the segment with higher priority. When the receiver successfully receives n packets out of N packets, the segments encoded with RS(N, k) codes can be correctly decoded, if $k \leq n$.

2.3 Error Protection for Aggregated Streams

For our distributed video conferencing system, we assume the outbound link of a video combiner node can transmit N packets with fixed length L symbols for every $1/F$ second. F is the frame rate of video source. We model the channel as a packet erasure channel in which each packet either arrives intact or is entirely lost. This can be achieved by inserting a sequence number in each sent packet and checking the sequence numbers at the receiver.

A logical step for the multi-stream aggregation is to first merge the base-layers from all streams and then to apply equal error protection to the merged base-layer stream. To apply unequal error protection to the merged FGS-layers, we examine two low overhead strategies as shown in Fig. 2.4. The shaded area indicates the source symbols and the white area indicates the RS code symbols.

The traditional strategy shown in Fig. 2.4(a) is a packet-based error protection strategy with time division multiplexing access (TDMA). As each user is assigned a set of packets and joint source/channel coding is performed within these assigned packets, users do not share packets. For user j, the video combiner needs to determine the number of packets, N_j, and select the RS code configuration for each segment belonging to user j. The second strategy, shown in Fig. 2.4(b), is segment based, and allows each user to transmit data in all available packets. For user j, we need to determine the number of segments assigned and the RS configuration of each segment. We refer to this new error protection strategy as PDMA. The overhead in the communication protocol introduced by these FEC strategies includes specifying the number of packets or the number of segments assigned to each stream, as well as specifying the source symbol assignment pattern for each seg-

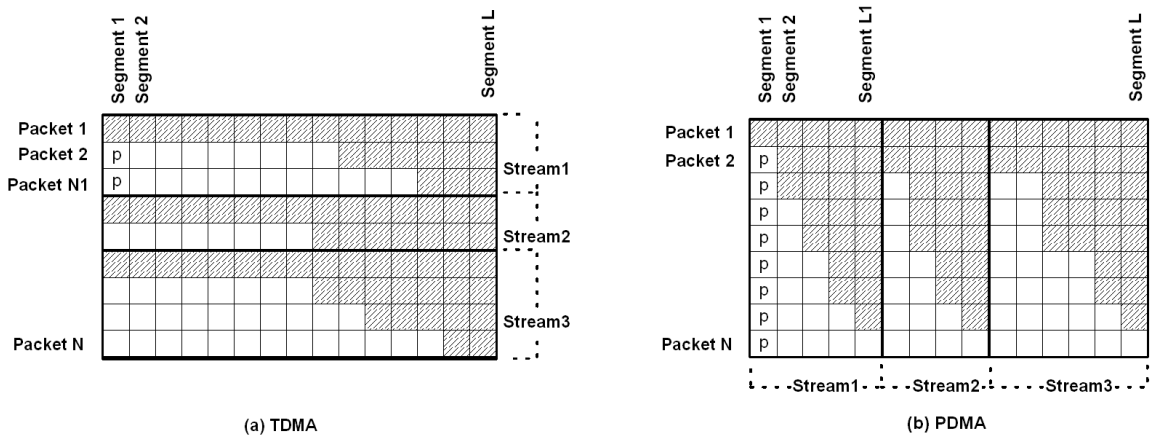


Figure 2.4: FEC strategies for multi-stream aggregation. (a) TDMA: Each stream is assigned a number of packets; (b) PDMA: Each stream is assigned a number of segments. White part indicates the RS code symbols. Shaded part indicates the source symbols.

ment. Further reduction of the communication overhead is possible by computing the source symbol assignment pattern of each segment on the receiver side that uses the same optimization algorithm based on the same R-D and channel information as the transmitter.

2.4 PDMA vs TDMA

Both TDMA-based and PDMA-based error protection approaches have low overhead for communicating the FEC pattern of merged video streams. Intuitively, because PDMA-based approach spreads error protection symbols to more packets than TDMA-based approach, PDMA-based approach may have better performance in error protection for packet-erasure channel. In addition, for the amount of error protection applied to the most important part of the source symbols, we observe

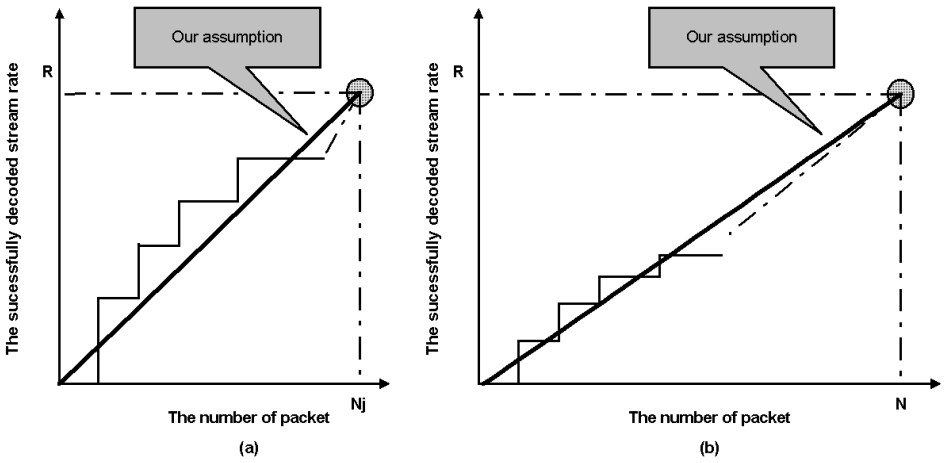


Figure 2.5: Rate-Packet function for one video stream. (a) TDMA-based FEC approach. (b) PDMA-based FEC approach.

that the actual protection applied in PDMA is more than that of TDMA as shown in Fig. 2.4 with "p" indicating the protection. Recent analytic studies in [61] have shown that the PDMA-based scheme has advantages over TDMA-based scheme in terms of expected throughput using equal error protection. In this section, we would like to compare PDMA-based scheme with TDMA-based scheme in terms of expected distortion using unequal error protection.

The expected distortion of delivered video streams is dependent on multiple factors, such as transmission channel condition, channel coding characteristics and video scene R-D characteristics. Therefore, it is quite involving to compare PDMA-based FEC approach vs TDMA-based FEC approach in terms of the expected video distortion. To shed the light in this issue, we perform analytical studies with some special-case assumptions for simplification.

We assume that there are J streams to be combined. The overall available

number of packets is N_F, and the packet length is L symbols. For the TDMA-based FEC approach, we assign L_j^a segments and N_{Fj}^a packets to the j^{th} stream. For PDMA-based FEC approach, we assign L_j^b segments and N_{Fj}^b packets to the j^{th} stream. Obviously, the following equations hold in our design:

$$\begin{aligned} &\textstyle\sum_{j=1}^{J} N_{Fj}^a = N \\ &L_j^a = L, \forall j \\ &\textstyle\sum_{j=1}^{J} L_j^b = L \\ &N_{Fj}^b = N, \forall j \end{aligned} \tag{2.1}$$

To facilitate our analysis, we make some assumptions regarding transmission channel condition, channel coding characteristics and video scene R-D characteristics.

First, we assume that the communication channel is a memoryless packet erasure channel. The packet successfully receive rate is denoted as p. The probability of successfully receiving n packets out of N_F is:

$$P_n^{N_F} = \begin{pmatrix} N_F \\ n \end{pmatrix} (p)^n (1-p)^{N_F - n}. \tag{2.2}$$

Second, we make an approximation of the channel coding characteristics. Without loss of generality, we focus on the distortion analysis of one video stream j in the merged video streams. Generally, if receiver receives more packets that contain the source symbols or error protection symbols of stream j, there are more successfully decoded symbols of stream j. The successfully decoded symbols as a function of received packets has a stair-case shape as shown in Fig. 2.5. We refer to this function as Rate-Packet (R-P) function. R-P functions are denoted as $r_{n,j}^a = \phi_j^a(n)$ and $r_{n,j}^b = \phi_j^b(n)$ for TDMA-based approach and PDMA-based approach. Here, $r_{n,j}^a$ and $r_{n,j}^b$ are successfully decoded symbol rate for TDMA-based approach and PDMA-based approach, respectively, and n is the number of received

packets. Since we use the same FEC method within one stream for comparing TDMA-based and PDMA-based approaches, we assume that there are same number of source plus FEC symbols assigned to j^{th} video stream in both approaches, i.e. $N^a_{Fj} \cdot L \approx N_F \cdot L^b_j$. Let $\frac{N^a_{Fj}}{N_F} = \frac{L^b_j}{L} = \alpha$, we have $0 < \alpha < 1$ for $J > 1$. In addition, we assume that there are same number of source symbols R_{Fj} assigned to this stream in both approaches. i.e. $R_{Fj} = r^a_{N^a_{Fj},j} = r^b_{N_F,j}$. Graceful R-P function with finer steps is generally desired because of the advantages in designing error protection and resource allocation strategy. To facilitate our analysis, we use linear function to approximate the R-P function. Since $\phi^a_j(0) = \phi^b_j(0) = 0$ and $\phi^a_j(N^a_{Fj}) = \phi^b_j(N_F) = R_{Fj}$, we use the following linear function to approximate R-P functions:

$$\begin{aligned} r^a_{n,j} &= \phi^a_j(n) = \beta^a_j \cdot n \\ r^b_{n,j} &= \phi^b_j(n) = \beta^b_j \cdot n \\ \beta^a_j &= \frac{R_{Fj}}{N^a_{Fj}} \\ \beta^b_j &= \frac{R_{Fj}}{N_F} \end{aligned} \tag{2.3}$$

Note that these linear functions may not be the linear approximation of R-P functions with minimum estimation error, they are special-case assumptions to simplify the analytical studies and give some insights to the comparison of TDMA-based FEC and PDMA-based FEC. We use β to denote the ratio of the R-P functions' slope, i.e., $\beta = \frac{\beta^a_j}{\beta^b_j} = \frac{R_{Fj}/N^a_{Fj}}{N_F/R_{Fj}} = \frac{N_F}{N^a_{Fj}} = \frac{1}{\alpha}$.

Third, we use an approximation of video scene R-D characteristics in our analysis. For multi-layer source codec such as MPEG-4 FGS, different FEC strategies are likely to be applied to different layer of source symbols as described in the previous section. For simplicity, our analysis is focus on one layer of source symbols. R-D characteristics of one layer of a compressed video frame can be approximated

to a summation of exponential terms [2, 22]:

$$D_j(r) = \sum_{s=1}^{s=S} C_{1j}^s \cdot e^{-C_{2j}^s \cdot r} + C_{3j}^s \tag{2.4}$$

Here, C_{1j}^s, C_{2j}^s and C_{3j}^s are all positive constants for a stream, and S is a pre-defined constant which is common for video frames.

Let ED_j^a and ED_j^b be the expected distortion of the j^{th} video stream for TDMA-base approach and PDMA-base approach, respectively. There could be a certain amount of symbols received for j^{th} stream other than the layer of symbols we are focused on, and we use $R_{0,j}$ to denote the received symbol rate of other layers. Let $D_j(r)$ denote the rate distortion function of the j^{th} video stream, the expectation of the received video stream distortion for TDMA-based FEC approach is:

$$ED_j^a = \sum_{n=0}^{N_{Fj}^a} P_n^{N_{Fj}^a} \cdot D_j(r_{n,j}^a + R_{0,j}) \tag{2.5}$$

The expectation of the received video stream distortion for PDMA-base FEC approach is:

$$ED_j^b = \sum_{n=0}^{N_F} P_n^{N_F} \cdot D_j(r_{n,j}^b + R_{0,j}) \tag{2.6}$$

Applying (2.3) to (2.5) and (2.6), we have:

$$\begin{aligned} ED_j^a &= \textstyle\sum_{n=0}^{N_{Fj}^a} P_n^{N_{F1}^a} \cdot D_j(\beta_j^a \cdot n + R_{0,j}) \\ ED_j^b &= \textstyle\sum_{n=0}^{N_F} P_n^{N_F} \cdot D_j(\beta_j^b \cdot n + R_{0,j}) \end{aligned} \tag{2.7}$$

Applying (2.4) to (2.7) and considering

$$\begin{aligned} &\textstyle\sum_{n=0}^{N} P_n^N = 1, \forall N \\ &\textstyle\sum_{n=0}^{N} e^{t \cdot n} P_n^N = [pe^t + (1-p)]^N, \forall N \end{aligned} \tag{2.8}$$

we have

$$\begin{aligned} &ED_j^a = \\ &\textstyle\sum_{s=1}^{s=S}\{C_{1j}^s \cdot e^{-C_{2j}^s \cdot R_{0,j}} \cdot [p \cdot e^{-C_{2j}^s \cdot \beta_j^a} + (1-p)]^{N_{Fj}^a} + C_{3j}^s\} \\ &ED_j^b = \\ &\textstyle\sum_{s=1}^{s=S}\{C_{1j}^s \cdot e^{-C_{2j}^s \cdot R_{0,j}} \cdot [p \cdot e^{-C_{2j}^s \cdot \beta_j^b} + (1-p)]^{N_F} + C_{3j}^s\} \end{aligned} \tag{2.9}$$

To compare ED_j^a to ED_j^b, we can evaluate the sign of $ED_j^a - ED_j^b$.

$$\begin{aligned} &ED_j^a - ED_j^b \\ &= \textstyle\sum_{s=1}^{s=S} M^s \cdot \{[p \cdot e^{-C_{2j}^s \cdot \beta_j^a} + (1-p)]^{N_{Fj}^a} \\ &-[p \cdot e^{-C_{2j}^s \cdot \beta_j^b} + (1-p)]^{N_F}\} \\ &= \textstyle\sum_{s=1}^{s=S} M^s \cdot \{[p \cdot e^{-C_{2j}^s \cdot \frac{\beta_j^b}{\alpha}} + (1-p)]^{\alpha \cdot N_F} \\ &-[p \cdot e^{-C_{2j}^s \cdot \beta_j^b} + (1-p)]^{N_F}\} \end{aligned} \tag{2.10}$$

Here, $M^s = C_{1j}^s \cdot e^{-C_{2j}^s \cdot R_{0,j}}$ is positive. The sign of $ED_j^a - ED_j^b$ is determined by the second term in (2.10). To evaluate its sign, let us take a look at a function as follows:

$$f(\alpha) = [p \cdot e^{-C_{2j}^s \cdot \frac{\beta_j^b}{\alpha}} + (1-p)]^{\alpha \cdot N_F} \tag{2.11}$$

Applying (2.11) to (2.10), we get:

$$ED_j^a - ED_j^b = \sum_{s=1}^{s=S} M^s \cdot (f(\alpha) - f(1)) \tag{2.12}$$

We can prove that $\frac{df(\alpha)}{d\alpha} < 0$ when $0 < \alpha \leq 1$, so that the second term in (2.12) is positive. Overall, we get $ED_j^a - ED_j^b > 0$, i.e. $ED_j^a > ED_j^b$.

In conclusion, the PDMA-based FEC approach may achieve lower expected distortion with the previously described transmission channel condition, channel coding characteristics and video scene R-D characteristics assumptions.

2.5 Chapter Summary

In this chapter, we propose a distributed multi-point video conferencing system over packet erasure channels. For this video conferencing system, we propose TDMA-based and PDMA-based error protection schemes for multi-stream aggregation that explores the multi-stream heterogeneity. Based on analytical studies, PDMA-based error protection scheme has superior performance in terms of delivered visual quality.

Chapter 3

Multi-Stream Joint Error Protection

To realize the multi-point video conferencing system proposed in the previous chapter, this chapter presents the formulation of PDMA-based error protection operation in each video combiner as an optimization problem. The TDMA-based approach can be formulated and solved in the same way by substituting the segment number with packet number in the problem formulation. Considering the challenge of supporting real-time multi-point video conferencing, we propose an iterative fast-search algorithm for PDMA-based allocation and provide simulation results to demonstrate the superior performance compared to traditional approaches.

This chapter is organized as follows. In Section 3.1, we formulate the error protection problem for proposed distributed video conferencing system as an optimization problem. In Section 3.2, an algorithm is then proposed to provide optimal solutions. Section 3.3 presents the experimental results. Discussions and chapter summary are presented in Section 3.4.

3.1 Problem Formulation

Suppose there are J video streams to be merged into N_F packets and there are L segments in each packet. The video combiner performs the packet merging for every incoming video frame. For simplicity, we omit the frame index from the notation in the subsequent discussions. In order to deploy the PDMA-based error protection, we need to determine the number of segments L_j to be allocated to the j^{th} stream and the number of RS protection symbols to be assigned to each segment.

To facilitate the discussion, let $a_{j,l} \in \{0,1\}$ be an indicator to represent whether segment l is allocated to user j. The overall segment-to-user assignment can be represented as $\mathbf{A}$, a $J \times L$ matrix with $[\mathbf{A}]_{j,l} = a_{j,l}$. In addition, we use $f_{i,l} \in \{0,1\}$ as an indicator to represent whether the number of source symbols assigned to segment l is greater than or equal to i. The overall source symbol-to-segment assignment can be represented as $\mathbf{F}$, a $N_F \times L$ matrix with $[\mathbf{F}]_{i,l} = f_{i,l}$.

Let $D_j(r)$ denote the distortion function of a video frame from j^{th} user when the receiving rate of FGS-layer source symbol is r. For simplicity, we assume that the base-layer source symbols of this frame can all be successfully decoded because of strong error protection. Suppose the receiver located in the next hop receives exactly n packets when the video combiner sends N_F packets, the reconstructed video quality in terms of distortion to the original video frame for user j can be represented as follows:

$$D_{j,n}(\mathbf{A},\mathbf{F}) = D_j(\sum_{l=1}^{L}\sum_{i=1}^{n} a_{jl}\bar{f}_{il}),\ where\ \bar{f}_{il} = \begin{cases} f_{il},\ if\ \sum_{i=1}^{N_F} f_{il} \le n \\ 0,\ if\ \sum_{i=1}^{N_F} f_{il} > n \end{cases} \tag{3.1}$$

The distortion reduction of correctly receiving one more correct packet after successfully receiving $n-1$ packets is $\Delta D_{j,n}(\mathbf{A},\mathbf{F}) = D_{j,n-1}(\mathbf{A},\mathbf{F}) - D_{j,n}(\mathbf{A},\mathbf{F})$. Let

p_c be the packet loss rate of the channel from a video combiner to the next hop and $P_c(N_F, n)$ be the probability that the receiver receives at least n packets successfully when the transmitter sends N_F packets. We have:

$$P_c(N_F, n) = \sum_{\alpha=n}^{N_F} \begin{pmatrix} N_F \\ \alpha \end{pmatrix} (1-p_c)^{\alpha} (p_c)^{N_F-\alpha}. \tag{3.2}$$

Let $D_{j,0}(\mathbf{A}, \mathbf{F})$ denote the distortion of a video frame when there is no FGS-layer packet received. The expected distortion of transmitting N_F packets of user j using segment assignment $\mathbf{A}$ and RS source symbol assignment $\mathbf{F}$ can be expressed as:

$$ED_j(\mathbf{A}, \mathbf{F}) = D_{j,0}(\mathbf{A}, \mathbf{F}) - \sum_{n=1}^{N_F} P_c(N_F, n) \Delta D_{j,n}(\mathbf{A}, \mathbf{F}). \tag{3.3}$$

To provide good video quality to all users as well as fairness across users, we formulate the problem as the following min-max optimization problem:

$$\min_{\mathbf{A},\mathbf{F}} (\max_{j} \ w_j \cdot ED_j(\mathbf{A}, \mathbf{F})) \tag{3.4}$$

subject to

$$\begin{cases} \sum_{l=1}^{L} a_{j,l} = L_j, \sum_{j=1}^{J} L_j = L; \\ \sum_{j=1}^{J} a_{j,l} = 1, \forall l; \\ f_{i,l} \geq f_{i+1,l}, f_{i,l} \in \{0,1\}, \forall i, l; \\ \sum_{i=1}^{N_F} f_{i,l} \leq \sum_{i=1}^{N_F} f_{i,l+1}, \text{ if } \exists j, \exists l, s.t.\ a_{j,l} = a_{j,l+1} = 1; \end{cases}$$

Here, w_j is the quality weight factor. By setting different w_j values for different video streams, our scheme can achieve differentiated quality among the received video streams.

In the problem formulation (3.4), the first constraint restricts that there are a total of L segments to be assigned to J streams. The second constraint is the segment assignment constraint for $\mathbf{A}$, requiring that each segment can be assigned

to only one video stream. The third and forth constraints are the source symbol assignment constraints for $\mathbf{F}$. For unequal error protection, we apply stronger RS codes for data with higher priority in error protection, i.e., $\sum_{i=1}^{N_F} f_{i,l} \leq \sum_{i=1}^{N_F} f_{i,l+1}$, if segments l and $l+1$ are allocated to the same video stream. The solution to (3.4) gives the optimal $\mathbf{A}$ and $\mathbf{F}$, which determine the information of the number of segments allocated to each stream and the number of source symbols assigned to each segment, respectively.

3.2 Proposed Algorithm

As mentioned in Section 2.2, MPEG-4 FGS is a two-layer video codec and each layer has different importance for error protection. We adopt different error protection schemes for each layer. We denote the outbound bandwidth of a video combiner as B_c bits per second, then the maximum number of packets for a video frame that the video combiner can send to the next hop is $N = \lfloor B_c/(sFL) \rfloor$, where s is the number of bits per symbol and F is the number of video frames per second.

3.2.1 Base-Layer Bandwidth Allocation and Error Protection

Strong equal error protection is applied to the base-layer source symbols. The encoder generates a base-layer at a low bit-rate R_j^b for a video frame from user j using a large quantization step, in order to ensure that the bandwidth is enough to transmit the base-layer and its protection symbols. We aggregate all users' base-layer data into $N_B^S = \lceil \sum_{j=1}^{J} R_j^b/(sL) \rceil$ source packets. It has been shown that if the packet loss rate (PLR) after FEC decoding can be kept below a threshold, $\text{PLR}^B = 10^{-3}$, the distortion caused by the channel error is negligible for MPEG-4 codec

[24]. We can find the minimum number of FEC packets, N_B^P, to achieve the desired PLR threshold: $P_c(N_B^S + N_B^P, N_B^S) \geq (1 - \text{PLR}^B)^{N_B^S}$. The overall number of packets for the base-layer is $N_B = N_B^S + N_B^P$.

3.2.2 FGS-Layer Resource Allocation via PDMA Bi-Section Search

A PDMA-based unequal error protection is applied to the FGS-layer source symbols. After N_B packets are assigned to the base-layer, there are $N_F = N - N_B$ packets to be assigned to the FGS-layer. We propose a bi-section search algorithm to solve this PDMA-based unequal error protection problem as formulated in (3.4).

Step 1: Obtain the segment-to-expected-distortion curve.

For a video frame $\bar{j}$, given $L_{\bar{j}}$ segments, we can obtain the corresponding minimum expected distortion based on MD-FEC [52, 59]. Aiming at minimizing the expected distortion of a single video stream, the original MD-FEC scheme provides the solution to the following problem of a single video stream:

$$ED^{min} \triangleq \min_{\mathbf{A},\mathbf{F}}(ED_{\bar{j}}(\mathbf{A}, \mathbf{F})) \tag{3.5}$$

subject to

$$\begin{cases} \sum_{l=1}^{L} a_{\bar{j},l} = L_{\bar{j}}; \\ f_{i,l} \geq f_{i+1,l}, f_{i,l} \in \{0,1\}, \forall i, l; \\ \sum_{i=1}^{N_F} f_{i,l} \leq \sum_{i=1}^{N_F} f_{i,l+1}, \text{ if } \exists l, s.t.\ a_{j,l} = a_{j,l+1} = 1; \end{cases}$$

Here, the matrix components $a_{j,l}$ and $f_{i,l}$ have been defined in Sec. 3.1. $a_{j,l} = 1$ for $l \leq L_{\bar{j}}$ and $j = \bar{j}$, $a_{j,l} = 0$ otherwise. $ED_{\bar{j}}$ is the expected distortion of stream $\bar{j}$. The objective function is to minimize the expected distortion of a single video stream subject to the bandwidth limitation for this single stream and the constraints for assignment of RS codes. These constraints are similar to those constraints in problem (3.4) except only a single video stream is involved.

For each j^{th} video stream ($j \in \{1, 2, \cdots J\}$) and the available number of segments L_j ($L_j \in \{1, 2, \cdots L\}$), the fast algorithm in [59] that has moderate computational complexity can be used to find the minimum expected distortion ED^{min} for problem (3.5). We denote this minimum expected distortion ED^{min} of j^{th} stream as S_j for simplicity. For a total of L segments, there are a total of L minimum expected distortion values S_j. We denote these values and their corresponding segment numbers as a row vector $[S_j, L_j]$. We refer to each of such vectors as a "Segment-Distortion (S-D) pair" and these pairs form a set $\mathcal{M}$ for a total of J video streams. Because the expected distortion S_j can be reduced if more segments are assigned to this stream, S_j is non-increasing with respect to L_j for a single video stream. For j^{th} stream, we can then piecewise interpolate S_j with respect to L_j to obtain a segment-to-expected-distortion curve $S_j(L_j)$. We refer to this curve as a S-D curve.

Step 2: Perform bi-section search.

We can show that the solution of problem (3.4) lies in the set $\mathcal{M}$ by the following two steps.

First, the constraints in (3.5) form a subset of the constraints in (3.4). If there exists $(\bar{\mathbf{A}}, \bar{\mathbf{F}})$ as the solution for (3.4), it should satisfy the constraints in (3.5).

Next, we show the solution of (3.4) achieves the minimum expected distortion defined in (3.5) and the corresponding L_j segments assigned to a single video stream. Let $\bar{j} = \arg\max_j(w_j \cdot ED_j(\bar{\mathbf{A}}, \bar{\mathbf{F}}))$ and $\bar{ED}$ be the achieved minimum expected distortion of (3.5) for stream $\bar{j}$, we get $\bar{ED} = \min_{\mathbf{A},\mathbf{F}}(ED_{\bar{j}}(\mathbf{A}, \mathbf{F})) = ED_{\bar{j}}(\bar{\mathbf{A}}, \bar{\mathbf{F}})$. Obviously, $(\bar{\mathbf{A}}, \bar{\mathbf{F}})$ achieves the minimum expected distortion defined as objective function of (3.5) for $\bar{j}^{th}$ stream.

In summary, if there exists $(\bar{\mathbf{A}}, \bar{\mathbf{F}})$ as the solution for (3.4), it also satisfies constraints of (3.5) and achieves its objective function. Therefore, the minimum

expected distortion achieved by $(\bar{\mathbf{A}}, \bar{\mathbf{F}})$ and its segment assignment lie in $\mathcal{M}$, i.e., the S-D pair set $\mathcal{M}$ of (3.5).

Since the minimum expected distortion and the corresponding segment assignment of (3.4) lie in $\mathcal{M}$, we can find the optimal solution of (3.4) by searching minimum expected distortion achieved by S-D pairs in $\mathcal{M}$, i.e., solving the following problem:

$$\min_{\{L_j\}}(\max_{j} \ (w_j \cdot S_j(L_j))), \quad \textit{subject to} \ \sum_{j=1}^{J} L_j \leq L. \tag{3.6}$$

Each S-D curve $S_j(L_j)$ is non-increasing with respect to the segment number L_j. To solve (3.6), we perform a bi-section search to obtain the optimal S-D pair for each stream. Given a distortion value, the bi-section search algorithm calculates the required number of segments of each stream. If the total number of required segments is higher than the number of overall available segments, L, the distortion value can be increased at the next iteration, and vice versa. This search procedure stops when the total number of required segments is equal to L. As the solution L_j^b by bi-section search algorithm that achieves the min-max distortion may not be an integer, we perform a small amount of search on the L_j^b round-up by finding $min_{L_j \in \lceil L_j^b \rceil, \lfloor L_j^b \rfloor\}, j \in \{1,2,\cdots J\}}(max_j(w_j \cdot S_j(L_j)))$, subject to $\sum_{j=1}^{J} L_j = L$.

Once the optimal S-D pairs are obtained from the bi-section search, a complete solution to (3.4) then includes the corresponding segment assignment and the optimal source symbol assignment within each stream provided by the original MD-FEC scheme [59].

3.2.3 FGS-Layer PDMA

The real-time interactive nature of video conferencing requires the video streams to be delivered promptly after being received by each video aggregation combiner

node. Error protection schemes with high computational complexity can cause undesirable delay for video delivery in each video aggregation combiner node. This motivates us to examine the computationally expensive part of the algorithm and investigate how to reduce the computational complexity of our proposed PDMA bi-section search algorithm. Based on our experimental results, one of the most computationally expensive parts is to obtain $L \times J$ S-D pairs of (3.5) in the first step of bi-section search. The overall computational complexity of this bottleneck part is $O(JNL^2)$, because it involves performing a search for RS code configuration [59] for $L \times J$ times and computational complexity of each search is $O(NL_j)$.

We now propose a fast algorithm to reduce the computational complexity of PDMA bi-section search by reducing the number of times to perform the RS code search described previously. Instead of calculating a total of $L \times J$ S-D pairs, this fast algorithm chooses a good initial segment-partition point, and then exploits an iterative technique.

Step 1: Initialization. It is critical to start with an initial point, $\{L_j^{(0)}\}$, which is close to the optimal segment-partition. We determine the initial point by considering an error-free channel and reformulate the problem (3.6) as:

$$\min_{\{L_j^{(0)}\}} (\max_j \ (w_j \cdot D_j(L_j^{(0)}))), \quad \textit{subject to} \ \sum_{j=1}^{J} L_j^{(0)} \leq L. \tag{3.7}$$

Here, the expected distortion S_j in (3.6) caused by channel errors and rate shaping is substituted by the deterministic distortion D_j caused by rate shaping only. As described in Section 2.2, R-D curves are made available to CN node by embedding them in video streams. Similar to bi-section search on S-D curves described earlier in this section, we can use bi-section search algorithm on R-D curves to obtain solution of (3.7), and use this solution, $\{L_j^{(0)}\}$, as the initial segment partition of

PDMA. Since the R-D characteristics of FGS-layer in one scene should not change dramatically from one frame to the following frame, we can also use the optimal segment assignment of the previous frame as an initial point of current frame PDMA.

After determining the initial point, we can obtain the expected distortion, $S_j(L_j^{(0)})$, for each user.

Step 2: Coarse Search. In this step, we determine the searching direction toward the optimal segment assignment. We take one segment from the video stream that has the smallest expected distortion and give one more segment to the video stream that has the largest expected distortion. Assume that k iterations have been performed, and the streams with the largest and the smallest distortion are j_{max} and j_{min} respectively:

$$\begin{aligned} j_{max} &= \arg\max_j(w_j \cdot S_j(L_j^{(k)})), \\ j_{min} &= \arg\min_j(w_j \cdot S_j(L_j^{(k)})), \\ ED_{max}^{(k)} &= \max_{\{\forall j\}}(w_j \cdot S_j(L_j^{(k)})). \end{aligned} \tag{3.8}$$

We take one segment away from user j_{min} and allocate it to user j_{max}:

$$\begin{aligned} L_{j_{max}}^{(k+1)} &= L_{j_{max}}^{(k)} + 1, \\ L_{j_{min}}^{(k+1)} &= L_{j_{min}}^{(k)} - 1, \\ L_j^{(k+1)} &= L_j^{(k)}, \forall j \neq j_{max}, j \neq j_{min} \end{aligned} \tag{3.9}$$

The corresponding expected distortion for the two users who are involved in segment re-allocation, namely, $S_{j_{max}}(L_{j_{max}}^{(k+1)})$ and $S_{j_{min}}(L_{j_{min}}^{(k+1)})$, are also updated. If the expected distortion difference between $S_{j_{max}}(L_{j_{max}}^{(k)})$ and $S_{j_{max}}(L_{j_{max}}^{(k+1)})$ is smaller than δ , we exclude user j_{max} in the next iteration, since the expected distortion of stream j_{max} would not be significantly improved by assigning more segments to it. $\delta = 10^{-3}$ is used in our experiments. This exclusion process allows segments to be assigned to video streams that show more significant improvement in expected distortion,

and the overall average distortion of all streams can be improved with negligible increase in the min-max distortion across streams. The above coarse search procedure is repeated until $ED_{max}^{(k+1)} \geq ED_{max}^{(k)}$.

Step 3: Finer Search. In coarse search, we only examine the video streams with the minimum and maximum expected distortion, and change segment assignments to these two video streams. As it is possible to further reduce the maximum expected distortion by examining other video streams, we perform a round of finer iterative search that considers all of video streams to obtain the min-max expected distortion.

Let $ED_{max}^{(k-1)}$ be the maximum distortion and denote the corresponding stream as j_{max}. In the k^{th} iteration of finer search, we perform $J-1$ trials by taking one segment from stream j and assigning it to stream j_{max}, $j \in \{1, 2, ...J\}$ and $j \neq j_{max}$. If the maximum distortion by segment reassignment is smaller than $ED_{max}^{(k-1)}$, it suggests that there exists a better solution than the result in the $(k-1)^{th}$ iteration. We can use this segment reassignment as a new start-point, and move to next round of coarse search. Otherwise, if the maximum distortion in no trial is smaller than $ED_{max}^{(k-1)}$, search is complete and $\{L_j^{(k-1)}\}$ is the optimal segment assignment.

The computational complexity of this fast algorithm is $O(TNL) + O(T'JNL)$, where T is the number of search iterations in step two and T' is the number of search iterations in step three. The sum of both iterations are typically less than 15 in our experiments. The computational complexity of PDMA fast algorithm is significantly reduced from the PDMA bi-section search when the total segment number L is much larger than 15.). This PDMA fast algorithm has been implemented in C/C++ with a moderate amount of optimization and experimented with a PC with Pentium Dual Core 2.8G Hz CPU and 2G RAM. The average computation time for merging 15

QCIF video frames is 25ms. In other words, in the experiment of merging 15 QCIF video frames, the PDMA fast algorithm can be real-time complemented within the frame refreshing time (around 33ms for 30 fps).

The optimal solution to (3.4) is not unique and there may exist several sets of solutions with the same min-max distortion. We can show by contradiction that there does not exist a better solution that can achieve smaller min-max distortion across streams than the one achieved by this PDMA fast algorithm.

Let the set of optimal solution provided by PDMA fast algorithm is $\{L_j^{opt}\}$. The frame from video stream $\bar{j}$ has the maximum distortion $\bar{S_1} = w_{\bar{j}} \cdot S_{\bar{j}}(L_{\bar{j}}^{opt})$. According to the step three of the proposed algorithm, for every video stream j other than $\bar{j}$, the expected distortion resulted by assigning one fewer segment than $\{L_j^{opt}\}$ can be larger than the optimal expected distortion:

$$w_j \cdot S_j(L_j^{opt} - 1) \geq w_{\bar{j}} \cdot S_j(L_{\bar{j}}^{opt}) \ \forall j \neq \bar{j}. \tag{3.10}$$

Suppose $\{L_j^{\star}\}$ is another set of solution that can provide smaller maximum distortion, denoted as $\bar{S_2}$. The total number of segments in both sets of optimal solutions should be equal to the maximum number of segments, L:

$$\sum_{j=1}^{J} L_j^{opt} = \sum_{j=1}^{J} L_j^{\star} = L. \tag{3.11}$$

Rearranging (3.11), we get

$$\sum_{j=1}^{J} (L_j^{opt} - L_j^{\star}) = 0. \tag{3.12}$$

Since $\bar{S_1} > \bar{S_2}$, we have $L_{\bar{j}}^{opt} < L_{\bar{j}}^{\star}$. So, based on (3.12), there exists at least some $\beta \in \{1, 2, ...J\}$, such that $L_{\beta}^{opt} > L_{\beta}^{\star}$. From (3.10), we have:

$$w_\beta \cdot S_\beta(L_\beta^{\star}) \geq w_\beta \cdot S_\beta(L_\beta^{opt} - 1) \geq w_{\bar{j}} \cdot S_j(L_{\bar{j}}^{opt}). \tag{3.13}$$

As revealed by (3.13), the maximum distortion from the segment assignment set $\{L_j^\star\}$ is not smaller than the maximum distortion achieved by $\{L_j^{opt}\}$. It contradicts the assumption that set $\{L_j^\star\}$ can provide smaller maximum expected distortion across streams. Therefore, there is no better solution to problem (3.4) that can achieve smaller min-max distortion across streams than the one achieved by PDMA fast algorithm.

3.3 Experimental Results

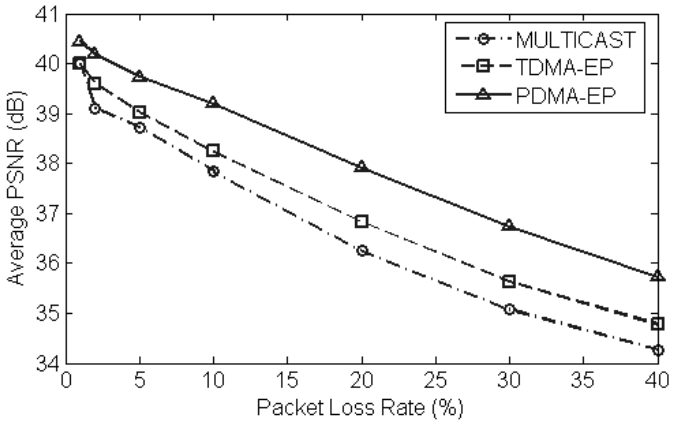
We evaluate the effectiveness of our proposed multi-stream PDMA error protection scheme (referred to as *PDMA-EP* in short) by comparing it to two alternative schemes. In the first alternative scheme referred to as *MULTICAST*, the bandwidth is divided evenly among streams. The second alternative scheme uses the same error protection approach that we proposed for the base-layer, but exploits the TDMA pattern as shown in Fig. 2.4(a). We refer to this scheme as TDMA Error Protection, or *TDMA-EP* in short. We first show the performance characteristics of each error protection scheme by a single-hop set-up, and then present the result for a multi-point video conferencing system in a memoryless packet erasure channel.

3.3.1 Single-Hop Experimental Results

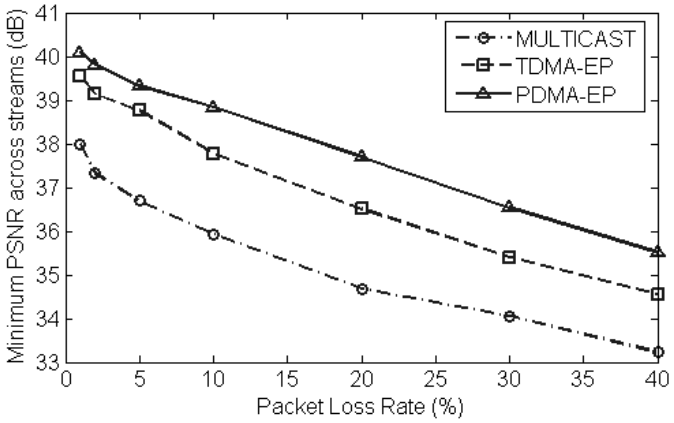
In our single-hop experiments, four video streams are aggregated by a video combiner. These four streams are "Suzie", "Akiyo", "Claire" and "Grandma", and referred to as stream 1 to 4, respectively. Each stream is encoded into 30 frames per GOP, and each GOP is led by one I frame followed by 29 P frames. The base-layer of each stream is encoded with quantization parameter $Q = 30$. There are 8 bits

per RS symbol and the encoded bit-stream packet size is 128 symbols(bytes). To simplify the simulation, in our experiments, the bandwidth is allocated to the source and FEC symbols only. Communication protocol headers are not included in our simulation. Without loss of generality, we examine the case of consistent quality among all users by setting all w_j at the same value. Single video stream has frame-to-frame data rate fluctuation due to different encoding modes, i.e., intra-mode for I frame and predictive-mode for P frame. If we merge I frames from multiple streams together, followed by the merge of P frames from multiple streams, the frame-to-frame data rate fluctuation can be more intense than single video stream. To avoid this tremendous data rate fluctuation, I, P frames from different streams should be interleaved before merging. In our experiments, the interleaving pattern is that the j^{th} stream sends I frame at the time of $j \cdot \triangle$, where $\triangle = 1/30$ second.

We first evaluate the performance of *PDMA-EP*, *TDMA-EP* and *MULTI-CAST* for channels with varying packet loss rate and the same bandwidth (4.2 Mbps). Fig. 3.1 shows the video quality results of aggregating 150 frames from stream 1 through stream 4, where the left figure shows the average PSNR of all streams and the right figure shows the minimum PSNR across all streams. The PSNR results are averaged over frames and repeated 200 test runs. We can see from these figures that our proposed *PDMA-EP* scheme consistently outperforms the two alternatives. At low packet loss rate, the cause of distortion is dominated by rate shaping for coping with bandwidth limitation, so both *PDMA-EP* and *TDMA-EP* achieve moderate video quality gain over *MULTICAST*. Although the long term (averaged) packet loss rate for communication networks is usually small, the channel condition can be very dynamic. Packet loss rate of 10% or higher is not rare over a short period of time because of network congestion or extensive noise/interference.

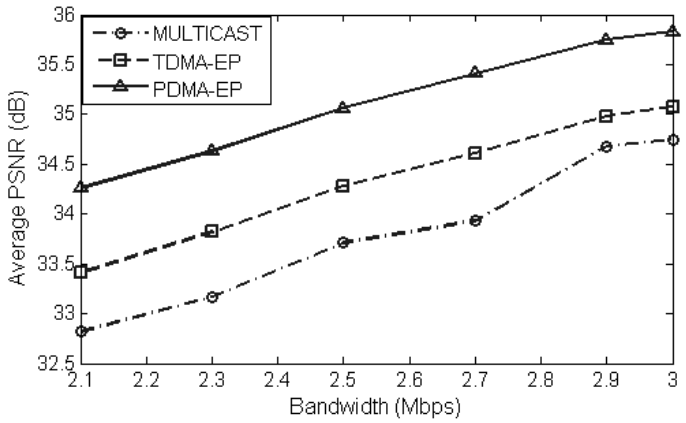


(a)

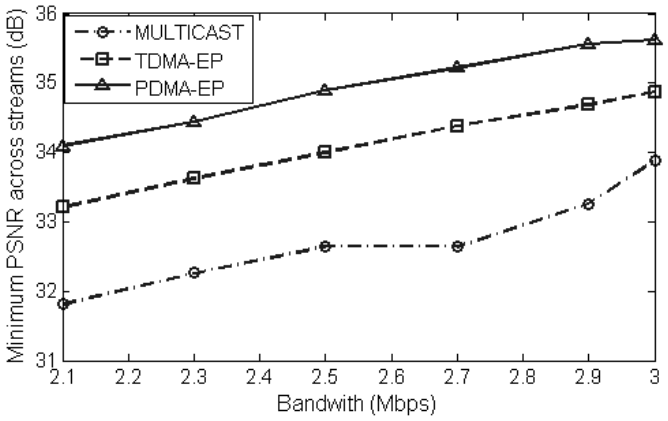


(b)

Figure 3.1: Schemes performance comparison with 4.2 Mbps bandwidth: Aggregating 150 frames from "Suzie", "Akiyo", "Claire" and "Grandma", respectively. PDMA-EP: Packet division multiplexing access error protection; TDMA-EP: Time division multiplexing access error protection; MULTICAST: Dividing the bandwidth evenly among the streams.



(a)



(b)

Figure 3.2: Schemes performance comparison with 20% packet loss rate: Aggregating 150 frames from 'Suzie", "Akiyo", "Claire" and "Grandma", respectively. PDMA-EP: Packet division multiplexing access error protection; TDMA-EP: Time division multiplexing access error protection; MULTICAST: Dividing the bandwidth evenly among the streams.

When the packet loss rate becomes larger, the gain tends to be more significant as error protection becomes a more effective factor on visual quality in transmission. In our experiment, when packet loss rate goes up to 40%, both *PDMA-EP* and *TDMA-EP* have up to 1.67 dB and 0.58 dB gain over *MULTICAST*, respectively. Comparing the performance of *PDMA-EP* and *TDMA-EP*, we observe that *PDMA-EP* has up to 1.10 dB gain over *TDMA-EP*. The performance gain of *PDMA-EP* over *PDMA-EP* is consistent with our analytical studies provided in Section 2.4. Another observation from Fig. 3.1 is that the difference between minimum PSNR and average PSNR for both *PDMA-EP* and *TDMA-EP* is small, only 0.31 dB for *PDMA-EP* and 0.35 dB for *TDMA-EP*. Such small difference indicates that these two schemes provide excellent fairness across multiple streams. In contrast, the difference of minimum PSNR and average PSNR for *MULTICAST* is as large as 1.62 dB. Because *MULTICAST* does not dynamically allocate resource to explore the multi-stream heterogeneity, it does not achieve good fairness of visual quality across streams.

In the second experiment, we fix the packet loss rate at 20% and evaluate the performance of these three schemes with a wide range of bandwidth. The results are shown in Fig. 3.2. Again, our proposed *PDMA-EP* scheme consistently outperforms the other two for variant bandwidth limitation. The average PSNR of *PDMA-EP* and *TDMA-EP* has up to 1.48 dB and 0.68 dB gain over *MULTICAST* in this experiment, respectively, and *PDMA-EP* has the average of 0.79 dB gain over *TDMA-EP*. When comparing the minimum PSNR across four streams, *PDMA-EP* and *TDMA-EP* have 2.22 dB and 1.38 dB average gain over *MULTICAST*, respectively, and *PDMA-EP* has an average of 0.84 dB gain over *TDMA-EP*. In terms of PSNR variation from stream to stream, we have similar observation as the

Table 3.1: The received video frame distortion with quality weighted factor.

	Suzie	Akiyo	Claire	Grandma
Weighed factor: w_j	0.1	0.2	0.3	0.4
Average MSE: ED_j	33.79	16.29	10.67	8.59
Average weighted MSE: $w_j \cdot ED_j$	3.379	3.258	3.201	3.436

first experiment, i.e., both *PDMA-EP* and *TDMA-EP* achieve significantly better fairness across streams than *MULTICAST*.

The third experiment is to demonstrate the capability of our scheme to provide desired differential visual quality among aggregated streams. In this experiment, the equal quality-weight-factor set-up is changed to be $w_1 = 0.1, w_2 = 0.2, w_3 = 0.3$, and $w_4 = 0.4$. The overall bandwidth is 3.0 Mbps, and packet loss rate is 10%. As shown in Table 3.1, the average distortion of each of the four streams differs in accordance to the specified weight factors. The stream with smaller weight factor has lower visual quality delivered as desired.

3.3.2 Experimental Results for A Multi-Point Video Conferencing

We implement a multi-point video conferencing system and carry out simulations to evaluate the performance of our proposed error protection scheme. The topology of a ten-user multi-point video conferencing is shown in Fig. 2.1. We again compare the performance of *PDMA-EP*, *TDMA-EP* and *MULTICAST*. The input test video streams of ten users are: Akiyo, Carphone, Claire, Foreman, Grandmother, Miss American, Mother & daughter, Salesman, Silent and Suzie. The set-

Table 3.2: The varying packet loss rate of video conferencing experiment for the topology shown in Fig. 2.1

Communication link	U_1	U_2	U_3	U_4	U_5	U_6
Packet loss rate	0.1	0.3	0.2	0.4	0.4	0.2
Communication link	U_7	U_8	U_9	U_{10}	C_{21}	C_{12}
Packet loss rate	0.1	0.4	0.2	0.3	0.5	0.2
Communication link	V_1	V_2	V_3	V_4	V_5	V_6
Packet loss rate	0.4	0.1	0.1	0.1	0.2	0.1
Communication link	V_7	V_8	V_9	V_{10}	C_{23}	C_{32}
Packet loss rate	0.5	0.2	0.4	0.1	0.4	0.3

tings of source encoding are the same as described in Section 3.3.1.

Because of the asymmetric data volume of the uplink and the downlink from user, the bandwidth of uplink is usually much smaller than the downlink bandwidth. In our experiments, the bandwidth of uplink for each user is set at 3 Mbps and the bandwidth of downlink for each user is 8.1 Mbps. The communication links between video combiner nodes are 9 Mbps. We perform experiments with a fixed packet loss rate at 10% and with varying packet loss rates at different communication links, respectively. In each experiment for video conferencing, one user receives 90 frames of video streams from each of the other users. The channel conditions for each of 24 communication links are listed in Table 3.2 and the quality weight factors are set to be equal for all aggregation communication links.

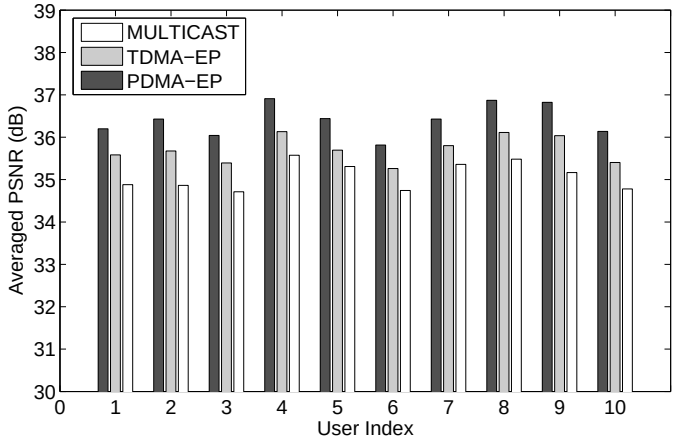
At 10% packet loss rate for all links, Fig. 3.3 shows the average PSNR and minimum PSNR across 9 received streams for each user averaged over 100 test runs and 90 frames. We can see from Fig. 3.3 that *PDMA-EP* can outperform *TDMA-EP* up to 0.8 dB, and outperform *MULTICAST* up to 1.7 dB. The average gain of *PDMA-EP* over *TDMA-EP* and *MULTICAST* is 0.7 dB and 1.3 dB, respectively.

For the experiment with varying packet loss rate, we vary the packet loss rate over a commonly seen range by generating random numbers that are uniformly distributed on the set $\{0.1, 0.2, 0.3, 0.4, 0.5\}$, and then using these random numbers as the packet loss rates of communication links. The set of the generated packet loss rates is listed in Table 3.2. Fig. 3.4 shows the average PSNR and minimum PSNR across 9 received streams for each user averaged over 100 test runs and 90 frames. In this experiment, *PDMA-EP* can outperform *TDMA-EP* by up to 0.97 dB, and outperform *MULTICAST* by up to 2.82 dB in PSNR. The average PSNR gain of *PDMA-EP* over *TDMA-EP* and *MULTICAST* is 0.76 dB and 1.64 dB,

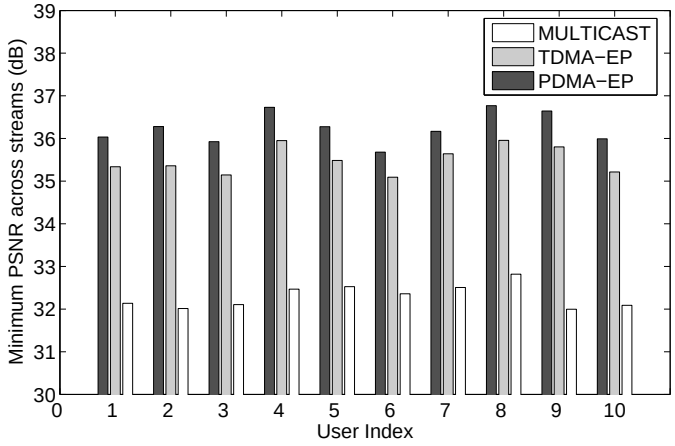
respectively. Compared to *MULTICAST*, both *PDMA-EP* and *TDMA-EP* have a larger gain in minimum PSNR among 9 received streams than the average PSNR. This is consistent with the single-hop results in Section 3.3.1. The experiments on other sets of randomly generated packet loss rates show similar results, i.e. *PDMA-EP* outperforms *TDMA-EP* and *MULTICAST* in terms of the visual quality of delivered video streams.

3.4 Chapter Summary

In this chapter, we formulate PDMA-based error protection for multi-stream aggregation to be a min-max optimization problem and propose an iterative search algorithm to achieve the optimal solution. Compared with TDMA-based and multicast-based error protection schemes, the proposed error protection scheme has up to more than 1 dB gain in terms of PSNR of delivered video streams.

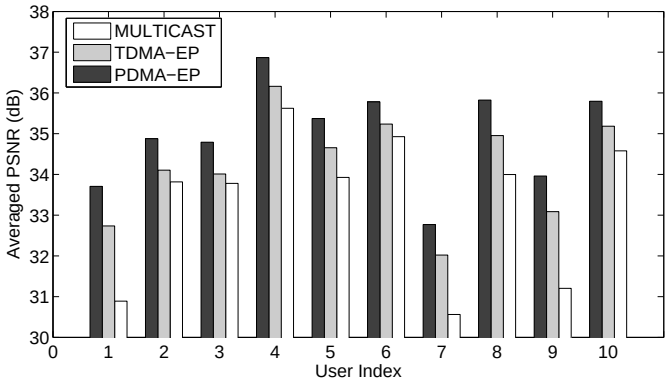


(a)

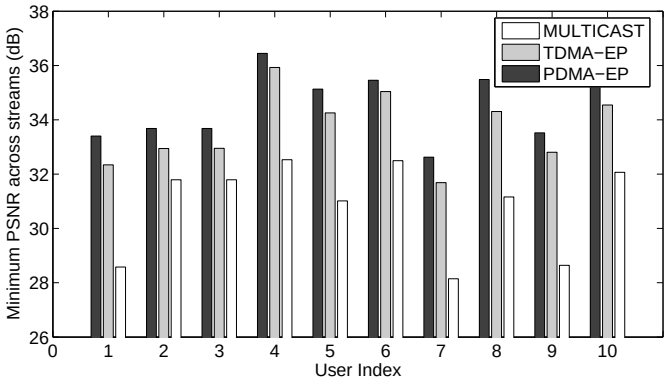


(b)

Figure 3.3: Schemes performance comparison with 10% packet loss rate for video conferencing session shown in Fig. 2.1. PDMA-EP: Packet division multiplexing access error protection; TDMA-EP: Time division multiplexing access error protection; MULTICAST: Dividing the bandwidth evenly among the streams. (a): Average PSNR. (b): minimum PSNR across received 9 streams.



(a)



(b)

Figure 3.4: Schemes performance comparison with varying packet loss rate for video conferencing session shown in Fig. 2.1. PDMA-EP: Packet division multiplexing access error protection; TDMA-EP: Time division multiplexing access error protection; MULTICAST: Dividing the bandwidth evenly among the streams. (a) Average PSNR; (b) minimum PSNR across received 9 streams.

Chapter 4

User Preference Heterogeneity

In the distributed multi-point video conferencing system examined in the previous chapter, each user receives all the other users' video streams. One important extension of this video conferencing application is that a user may have different interest or preference in incoming video streams [34, 46, 47, 48]. For example, a user may want to focus on conferee who is currently talking and want to receive this stream with higher quality than other streams in which participants have less activities. To extend our video conferencing system to support the varying preferences, we have defined the quality weight factor w_j in (3.4) to deliver video streams with differentiated quality. The problem of how to properly set w_j remains, especially at the intermediate communication links, in order to ensure good and fair video quality for all video conference attendees. This is because a video stream aggregated through the intermediate communication link may be delivered to users with heterogeneous quality preferences. In this chapter, we discuss how to derive w_j to support the preference of each participant for incoming video streams and solve the user-preference heterogeneity problem for video conferencing.

A traditional solution of user preference heterogeneity problem is to differ-

entiate delivered visual quality in the last hop [47, 48]. For a multi-point video conferencing system involving a mesh of multiple video combiners, this "last-mile" method would assign equal quality weight factors w_j in (3.4) for all $(CN_m - CN_n)$ transmissions, and only $(CN - UN_s)$ transmission would use the differentiated quality weight factors that are directly assigned by end users. An underlying assumption of the "last-mile" solution is that the intermediate communication link has much larger bandwidth than the last-mile communication link to the end user. Although this bandwidth assumption is generally true, a video conferencing session shares this bandwidth with thousands of other applications. The fast deployment of multimedia applications has made the data traffic in the intermediate communication links increasingly crowded. As a result, bandwidth reservation through Quality of Service (QoS), such as the Resource Reservation Protocol (RSVP), has been used in many deployments. The bandwidth of intermediate link available to a multimedia application may not be significantly larger than the last-mile link. This calls for investigating how to utilize the user preference information for multi-stream video aggregation over intermediate communication links. If the intermediate combiner node can take into account user preference when assigning the quality weight factors, the quality of delivered video streams can be improved, and better reflects the users' preference.

4.1 User Preference Heterogeneity Problem Formulation

As discussed above, for a $(CN - UN_s)$ link, the quality weight factor w_j for delivering the j^{th} stream to users can be set according to users' preference. For a $(CN_m - CN_n)$ link, delivering a video stream with quality lower than users' desired level may result in large maximum weighted distortion across streams defined in

(3.4). It is desirable to deliver a video stream with higher visual quality than the highest quality that all video conferencing users demand for that stream. More generally, we develop a mechanism for the end users in a conferencing system to first reach a consensus about the visual quality they demand, then aggregate multiple streams in an optimal way so that each stream can be delivered with quality close to or higher than the "consensus" quality.

To facilitate our discussion, we quantify the preference of user s for stream j as $\theta_{s,j}$. Note that $\theta_{s,s} = 0$, and $\theta_{s,j}$ is normalized so that $\sum_{j=1}^{J} \theta_{s,j} = 1$. Although we may be able to find optimal solution with packet loss rate taken account, considering the delay to transmit channel condition to video combiners, it is difficult for video combiners to be fully aware of the time varying channel condition. Since a higher throughput generally leads to higher visual quality for a video stream, by assuming the transmission channels are error-free, we simplify the problem of optimizing the visual quality of video conferencing to a problem of optimizing the throughput of video streams for end users over error-free channels. We now consider how to reach a consensus on user desired throughput from video combiner CN_m to CN_n. We denote Φ as the set containing the indices of user nodes whose streams are transmitted through video combiner CN_m to CN_n, and denote Ψ as the set containing the indices of user nodes whose received streams are transmitted through video combiner CN_m to CN_n. Let $J = |\Phi|$ be the number of streams to be merged and $S = |\Psi|$ be the number of users who receive those streams. Here, the operation $|\cdot|$ is to get the number of elements in a set. Let w_j be the normalized quality weight factor for stream j in Φ, and $\sum_{j\in\Phi} w_j = 1$. We use B_s^V to denote the bandwidth of a $(CN - UN_s)$ link V_s, and B^C to denote the bandwidth of a $(CN_m - CN_n)$ link C_{mn}. For user s in Ψ, the throughput of stream j along the $(CN - UN_s)$ link V_s are

considered as user s's preference. In other words, if the bandwidth utilization of link V_s is $\pi_s \in [0,1]$, the amount of data that user s receives for stream j is $\theta_{s,j} \cdot \pi_s \cdot B_s^V$. Based on the definition, the throughput of this stream j transmitted from CN_m to CN_n is $w_j \cdot B^C$. Therefore, considering the stream j for user $s \in \Psi$ in link V_s, the bandwidth utilization is:

$$\pi_s^j = \min\left\{1,\ \frac{w_j B^C}{\theta_{s,j} B_s^V}\right\}. \tag{4.1}$$

For the $(CN_m - CN_n)$ link C_{mn}, we would like to choose w_j such that the bandwidth utilization can be efficient for all of user s, $s \in \Psi$. For all the streams transmitted from CN_m to CN_n, we formulate a consensus problem to maximize the minimum bandwidth utilization for all users in Ψ:

$$\max_{\{w_j, \forall j \in \Phi\}} \min_{\{j \in \Phi;\ s \in \Psi\}} \pi_s^j, \quad \textit{subject to} \quad \sum_{j \in \Phi} w_j = 1; \tag{4.2}$$

4.2 Proposed Consensus Algorithm

The formulated optimization problem (4.2) can be solved by:

$$w_j = \frac{\overline{w_j}}{\sum_{j \in \Phi} \overline{w_j}}, \text{where } \overline{w_j} = \max\left\{\theta_{s,j} B_s^V,\ \forall s \in \Psi\right\} \tag{4.3}$$

To verify that (4.3) is the optimal solution to (4.2), we first rearrange the inner objective function in problem (4.2) as:

$$\min_{\{j \in \Phi;\ s \in \Psi\}} \pi_s^j = \min_{j \in \Phi}\{\min_{s \in \Psi} \pi_s^j\} = \min_{j \in \Phi}\{1, \frac{w_j B^C}{\theta_{s_j,j} B_{s_j}^V}\},\ \textit{where } s_j = \arg\max_{s \in \Psi}\{\theta_{s,j} B_s^V\}. \tag{4.4}$$

Bringing in the solution (4.3) into the minimization function in (4.4), we have:

$$\min_{\{j \in \Phi;\ s \in \Psi\}} \pi_s^j = \min\{1, \frac{B^C}{\sum_{j \in \Phi} \overline{w_j}}\}. \tag{4.5}$$

We consider the following two cases for the optimality of (4.3).

Case 1: $B^C < \sum_{j\in\Phi} \overline{w_j}$.

Suppose $\{w_j^\star,\ j \in \Phi\}$ is a set of optimal solution and $\pi_s^{\star\ j}$ is the corresponding bandwidth utilization. Then, $\{w_j^\star\}$ and $\{w_j\}$ should satisfy the constraint in problem (4.2):

$$\sum_{j\in\Phi} w_j^\star = \sum_{j\in\Phi} w_j = 1. \tag{4.6}$$

Rearrange (4.6), we have $\sum_{j\in\Phi}(w_j^\star - w_j) = 0$.

Suppose there exists a $w_\alpha^\star > w_\alpha$ for $\alpha \in \Phi$. Then there exists at least $w_\beta^\star < w_\beta$ for some $\beta \neq \alpha, \beta \in \Phi$. Then,

$$\frac{w_\beta^\star B^C}{\theta_{s_\beta\beta} B_{s_\beta}^V} < \frac{w_\beta B^C}{\theta_{s_\beta\beta} B_{s_\beta}^V} \tag{4.7}$$

From (4.4), (4.5), and (4.7), the minimum bandwidth utilization satisfies:

$$\min_{\{j\in\Phi;\ s\in\Psi\}} \pi_s^{\star\ j} < \min_{\{j\in\Phi;\ s\in\Psi\}} \pi_s^j. \tag{4.8}$$

This is a contradiction to $w_j^\star$ being the optimal solution. Therefore, $w_j = w_j^\star,\ \forall j \in \Phi$ is the optimal solution.

Case 2: $B^C \geq \sum_{j\in\Phi} \overline{w_j}$.

From (4.5), we have $\min_{\{j\in\Phi;\ s\in\Psi\}} \pi_s^j = 1$, which has achieved the maximum possible bandwidth utilization. Thus, $\{w_j\}$ is the optimal solution. Note that in this case, $\{w_j\}$ may not be the unique solution that provides maximum bandwidth utilization.

The consensus approach is suitable for the video conferencing system where the link to end user is relatively dedicated to a single user, such as the broadband DSL or cable modem links. In this case, the bandwidth of end user link is quite stable and can be known to all video combiner nodes. When end user bandwidth

B_s^V is the same from user to user, consensus algorithm (4.3) can be simplified to $w_j = \frac{\overline{w_j}}{\sum_{j\in\Phi}\overline{w_j}}$, where $\overline{w_j} = \max\{\theta_{s,j},\ \forall s \in \Psi\}$. This simplified approach can be used when the exact bandwidth for end users are not known, but approximately the same from user to user.

When the channel condition is taken into account, though (4.3) may not give optimal video quality anymore, our experimental results presented in the next section show that the proposed throughput-based optimal approach can still provide significant gain.

4.3 Experimental Results

To evaluate the effectiveness of our proposed consensus strategy, we first perform experiments on a two-CN-node case, then the experimental results on a multipoint video conferencing system with multiple hops are presented.

4.3.1 Experimental Results for A Two-CN-Node Case

The topology of this case is shown in Fig. 4.1, where UN4 and UN5 have significantly different preference for the video streams from UN1, UN2 and UN3. The preference values are listed in Table 4.1. We study the video combiner CN1 to illustrate the performance of our proposed consensus strategy. In this experiment, the packet loss rate is set to be 0.1 for all the communication links. The bandwidth is set to be 1.2 Mbps for U_1, U_2 and U_3, 2.1 Mbps for V_4 and V_5; the bandwidth limitation for C_{12} ranges from 2.1 Mbps to 4.5 Mbps. The three input frames for UN1, UN2 and UN3 are from "Suzie," "Akiyo" and "Claire," respectively. The traditional "last-mile" solution [47, 48] would set the quality weight factors as 1/3

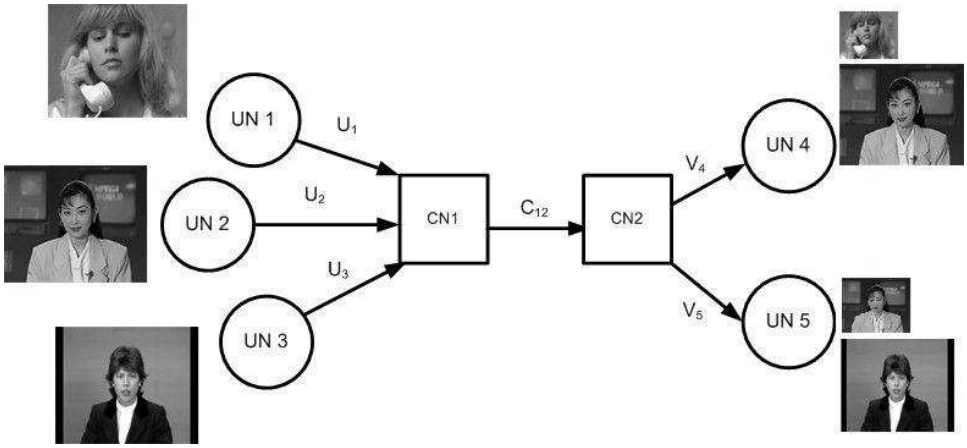


Figure 4.1: A two-video-combiner case: "Suzie","Akiyo" and "Claire" are aggregated from UN1, UN2 and UN3, to UN4 and UN5. UN4 is interested in receiving "Suzie" and "Akiyo" with the preference factor of 0.2 and 0.8; UN5 is interested in receiving "Akiyo" and "Claire" with the preference factor of 0.3 and 0.7 respectively.

for the aggregation link C_{12}, while the proposed consensus algorithm described in Section 4.2 gives the quality weight factors as listed in Table 4.1.

Fig. 4.2 shows the maximum of weighted distortion across streams received by each user, defined as the objective function to be minimized in (3.4). The results are averaged over 100 test runs. As we can see, by utilizing user preference information more effectively, the consensus strategy can provide much higher quality than the traditional "last-mile" solution. The improvement in the PSNR of the stream that has the maximum weighted distortion is up to 2.13 dB in this experiment. When the bandwidth of C_{12} is large enough so that the throughput of every video stream aggregated over C_{12} is larger than the throughput of the same video stream aggregated over V_4 and V_5, the performance of these two strategies become the same. This is

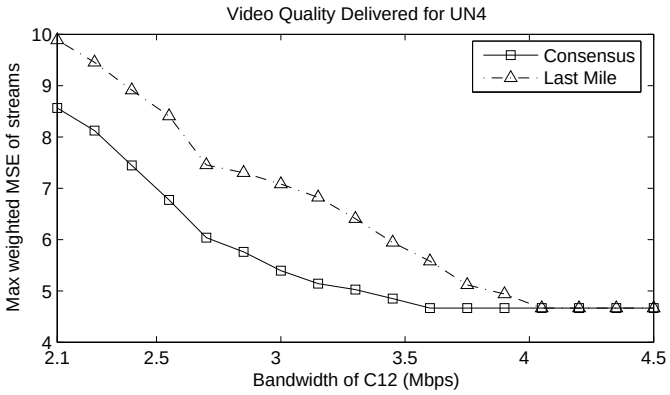
Table 4.1: Consensus strategy:"Consensus" column shows the preference value after the consensus procedure.

Video streams from user node UNx	User4's preference	User5's preference	Consensus preference
$UN1$	0.2	0.0	0.12
$UN2$	0.8	0.3	0.47
$UN3$	0.0	0.7	0.41

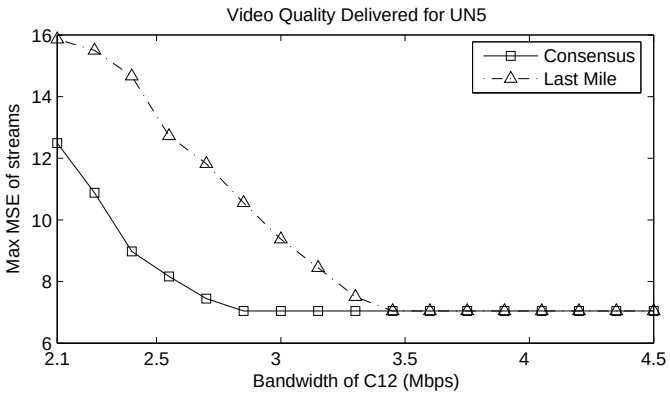
because that, in this case, the assumption of the "last-mile" solution becomes valid, i.e. the intermediate links have unlimited bandwidth [47, 48]. This observation is also consistent with our analysis in Section 4.2. The *Case 2* in Section 4.2 indicates that, when the bandwidth of C_{mn} is large enough, the bandwidth utilization is 100% and the optimal solution may not be unique.

4.3.2 Experimental Results for A Multi-Point Video Conferencing

We also perform an experiment on the ten-user multi-point video conferencing shown in Fig. 2.1 to compare the performance of proposed consensus algorithm to the "last-mile" solution. The encoded format of input streams is the same as described in Section 3.3.1. The bandwidth is set the same as the experiment in Section 3.3.2, and the packet loss rates are listed in Table 3.2. To simulate the varying user preference for different incoming video streams, we generate 90 random numbers which are uniformly distributed on the set $\{1, 2, 3, 4, 5\}$, where 1 indicates the lowest



(a) UN4



(b) UN5

Figure 4.2: Schemes performance comparison with packet loss rate 10%. The bandwidth is 2.1 Mbps for V_4 and V_5, and 1.2 Mbps for U_1, U_2 and U_3. The three input streams for $UN1$, $UN2$ and $UN3$ are "Suzie","Akiyo" and "Claire", respectively.

preference and 5 is the highest. These 90 random numbers are then grouped into 10 groups and normalized per group. We then use each group of numbers as the user preference for one user as shown in Table 4.2. Since a conference attendee's own stream does not need to be transmitted to himself/herself, the user preference values in the diagonal line of Table 4.2 are always zero.

In this experiment, one user receives 90 frames of video streams from all the other users. For both the consensus approach and "last-mile" approach, quality weight factor w_j in Equation (3.4) for V_i link is set to be the value in i^{th} row j^{th} column of Table 4.2. For intermediate links C_{12}, C_{21}, C_{23}, and C_{32}, our proposed consensus approach derives the quality weight factors as listed in Table 4.3, while for the "last-mile" approach, w_j is set to be the same value for all streams for C_{12}, C_{21}, C_{23}, and C_{32}. Fig. 4.3 shows the PSNR of the stream that has the maximum weighted distortion across 9 received streams for each user. It is averaged over repeated 100 test runs and 90 frames. We can see that the consensus approach either performs the same as "last-mile" approach or performs better. It has up to 2.64 dB gain improvement in this minimum PSNR over "last-mile" approach. For user 1-3 and user 8-10, consensus solution performs better and the average gain in terms of the PSNR of the stream that has the maximum weighted distortion is 1.24 dB. For user 4-7, the intermediate links that aggregate multiple streams towards them are C_{12} and C_{32}. Since C_{12} and C_{32} only aggregate three streams, these two links have relatively larger bandwidth per stream when we compare the bandwidth per stream to V_4, V_5, V_6 and V_7, where 9 streams are aggregated. Therefore, "last-mile" solution and consensus solution have the same performance for user 4-7.

Table 4.2: The user preference for incoming stream from user US_x to receiving node UN_x. The video conferencing topology is shown in Fig. 2.1.

Preference	US_1	US_2	US_3	US_4	US_5	US_6	US_7	US_8	US_9	US_{10}
UN_1	0	0.16	0.06	0.13	0.09	0.16	0.13	0.02	0.09	0.16
UN_2	0.08	0	0.12	0.12	0.15	0.12	0.08	0.03	0.15	0.15
UN_3	0.14	0.24	0	0.05	0.09	0.24	0.05	0.05	0.09	0.05
UN_4	0.15	0.08	0.04	0	0.04	0.15	0.12	0.18	0.12	0.12
UN_5	0.16	0.11	0.06	0.13	0	0.16	0.03	0.13	0.06	0.16
UN_6	0.14	0.09	0.04	0.16	0.14	0	0.04	0.16	0.09	0.14
UN_7	0.03	0.12	0.06	0.15	0.15	0.09	0	0.09	0.15	0.15
UN_8	0.19	0.07	0.07	0.14	0.14	0.07	0.11	0	0.14	0.07
UN_9	0.05	0.10	0.05	0.24	0.10	0.18	0.13	0.10	0	0.05
UN_{10}	0.16	0.08	0.03	0.16	0.16	0.06	0.06	0.16	0.13	0

Table 4.3: The consensus preference of link C_{mn} for incoming stream from user US_x

Preference	US_1	US_2	US_3	US_4	US_5	US_6	US_7	US_8	US_9	US_{10}
C_{12}	0.50	0.32	0.18	0	0	0	0	0	0	0
C_{21}	0	0	0	0.13	0.15	0.24	0.13	0.04	0.15	0.16
C_{23}	0.18	0.09	0.07	0.22	0.15	0.17	0.12	0	0	0
C_{32}	0	0	0	0	0	0	0	0.37	0.30	0.33

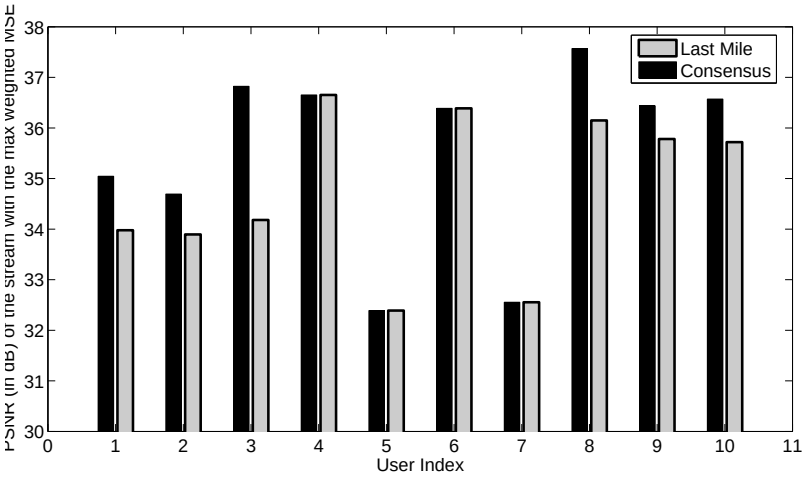


Figure 4.3: Schemes performance comparison in terms of average PSNR (dB) for video conferencing session shown in Fig. 2.1. The packet loss rates are shown in Table 3.2; The user preferences are listed in Table 4.2. Last Mile: Last-mile approach; Consensus: Proposed consensus approach.

4.4 Chapter Summary

To accommodate the user preference heterogeneity in a multi-point video conferencing system, in this chapter, we propose a distributed algorithm to reach consensus among all conferees. This algorithm is performed in each intermediate hop for aggregated streams, and thus the perceptual quality of delivered video streams can be improved compared to the "last-mile" solution.

Chapter 5

Multi-Hop Awareness

Distributed communication systems have advantages of flexible resource allocation and scalability for transmitting large number of multimedia streams. Multimedia services deployed over distributed systems have attracted a lot of research attention [8, 51, 75, 1] in recent years. The deployment of such multimedia systems often involves concurrently transmitting multiple video streams over sequential multiple hops. In a multi-hop environment, data payload arrives at an end user after being aggregated over a series of distributed nodes. In addition to performing the store-and-forwarding functions, most nodes in these systems are powerful computing devices. These nodes can implement complicated and intelligent tasks over all protocol layers, including source coding and decoding, channel coding and decoding, active routing, and quality-of-service (QoS) provisioning.

In a sequential multi-hop environment, a simple way to apply source and channel coding for multimedia data payload is to consider multiple hops from a transmitter to an end user as one communication link, and then process the data payload based on end-to-end channel condition. This approach may not achieve low visual distortion for delivered content because the bandwidth and error condition of

each hop may be different. In order to improve visual quality of delivered content, it has been suggested in the recent literature [56, 43, 31, 36] that applying hop-by-hop processing based on each hop's resource and channel condition can result in significant improvement in terms of the visual quality of transmitted streams. In [56], an overlay system is designed by partitioning end-to-end path into segments, and channel decoding and re-encoding is done in the intermediate nodes. An algorithm was proposed in [43] to reduce end-to-end delay of video stream transmission in a multi-hop wireless environment. In [31] and [36], algorithms were proposed to adjust rate allocation and channel coding in a coordinated fashion to minimize the visual distortion of a video stream transmitted over multiple parallel paths between two nodes.

Although hop-by-hop processing methods address some issues of video stream transmission in a multi-hop environment, transmitting multiple video streams over multiple unreliable hops faces more challenges. First, the data traffic in the intermediate communication hops has become increasingly crowded due to the fast-pace of multimedia application deployments. As a result, bandwidth reservation by Quality of Services (QoS) mechanism, such as Resource Reservation Protocol (RSVP), has been used in many deployments. Since a video application can have a limited bandwidth assigned in each communication hop, resource allocation across multiple video streams over multiple hops is essential for delivering multiple video streams with good visual quality. Considering that video streams transmitted over a hop may then be aggregated over multiple hops with heterogeneous channel conditions, in addition to the multi-stream and user preference heterogeneities discussed in the previous chapters, resource allocation needs to explore the bandwidth difference of multiple hops. Second, information loss caused by channel errors, such as the fading

or interference in wireless channel and congestion in wireline channel, creates another challenge for multi-hop multi-stream video applications. In real-time services, where the low delay is expected, FEC is a promising error-resilient technique. Supported by powerful distributed nodes, the idea of multi-hop FEC for single video stream transmission was explored in literature [56, 32, 21]. In these methods, each transmitter node solves FEC problem based on resource availability and channel reliability of one hop. There is no overall consideration of sequential multiple hop's resource and channel condition in these prior arts.

Given a limited amount of bandwidth, applying FEC and bandwidth allocation locally in one node may not achieve good visual quality for every aggregated video streams in a multi-hop multi-stream environment. As an example shown in Fig. 5.1, video streams 1 and 2 are first transmitted over hop 1, stream 1 is then transmitted over hop 2 to an end user, and stream 2 is transmitted over hop 3 to another end user. With the known channel condition of hop 1, i.e., bandwidth and channel error, the source and FEC symbols can be optimally assigned to achieve good visual quality for both video streams and maintain fairness across streams based on the algorithm proposed in Section 3.2. Similarly, with the known channel condition of the hop 2, the source and FEC symbols can be optimally assigned for stream 1 to achieve the minimum visual distortion of delivered content. Assuming hop 2 has severe channel loss, strong FEC should be applied to stream 1 achieve optimal visual quality. The strong FEC with limited bandwidth may result in that the number of source symbols of stream 1 demanded for optimal error protection assignment is less than the number of source symbols received. Certain number of source symbols of stream 1 delivered over hop 1 may be wasted and not aggregated over hop 2. In other words, even if these wasted source symbols are not successfully delivered over hop 1,

the visual quality of stream 1 should not be degraded for end user. However, after aggregated over hop 1, stream 2 is transmitted over a hop with relatively low channel loss. There are more source symbols of stream 2 demanded than received from hop 1. The video content loss of stream 2 over hop 1 cannot be recovered in hop 2, no matter how much bandwidth is available and how strong FEC is applied. In order to improve the visual quality of stream 2 delivered to end user, we should allocate more bandwidth to stream 2 in hop 1. In this way, visual quality of stream 2 delivered to the end user can be improved without degrading the visual quality of stream 1 delivered to the end user. This better strategy of bandwidth allocation requires hop 1 to be aware of the channel condition of hops 2 and 3. This observation motivates us to investigate the multi-hop awareness for multiple video streams' transmission. If we allocate resource and apply FEC to multiple video streams with multi-hop awareness, we should be able to improve the overall visual quality of the delivered video streams.

Based on multi-hop awareness, we propose a multi-hop multi-stream video aggregation scheme. This scheme searches for an optimal resource allocation and FEC configuration that provides good and fair visual quality to all video stream consumers. It explores the heterogeneity in the content characteristics, and in channel conditions of multiple hops. We also discuss practical issues of multi-hop awareness when applying the proposed scheme to real-time video applications. The simulation results show that our proposed scheme can outperform the scheme without multi-hop awareness in terms of the visual quality of delivered content.

The rest of this chapter is organized as follows. In Section 5.1, we formulate the multi-hop multi-stream video aggregation problem as an optimization problem. In Section 5.2, we propose an algorithm to solve this problem. Experimental results

are shown in Section 5.3 and conclusions are drawn in Section 5.4.

5.1 Multi-Hop Multi-Stream Aggregation Problem Formulation

Using the same building blocks as described in Section 2.2 of Chapter 2 [15], in this section, we formulate the multi-hop multi-stream video aggregation over packet erasure channel as an optimization problem. The problem formulation explores the heterogeneity in video stream characteristics and channel conditions, as well as multi-hop awareness. Although multi-hop multi-stream video aggregation can be applied in many multimedia applications, the example application we focus on is the proposed distributed multi-point video conferencing system as illustrated in Fig. 2.1.

To achieve good visual quality of a single video stream delivered in a multi-hop environment, each transmitter node should perform FEC and rate shaping in a distributed manner [11, 56, 43, 31]. Similarly, multi-stream aggregation scheme can be performed in each transmitter node in the multi-hop multi-stream environment to obtain good visual quality. Multiple video streams are merged frame-by-frame as described in the previous chapters. For simplicity, we omit the frame index from the notation in latter discussions.

Suppose there are J video streams to be transmitted over K communication hops to U end users with fixed packet length L, for k^{th} hop, we first use equal error protection algorithm described in Section 3.2 to apply error protection to base-layers of multiple video streams. Then, there are N packets and L segments left to be used for FGS-layers. In order to deploy the PDMA-based error protection, we need to determine the number of segments L_j to be allocated to the j^{th} stream in the k^{th} hop, and the number of RS protection symbols to be assigned to each

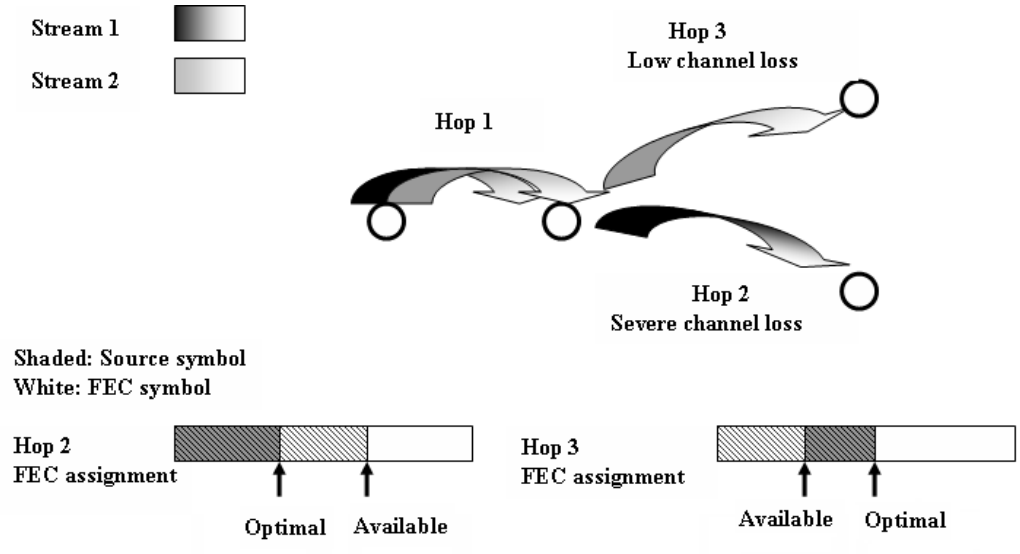


Figure 5.1: A simple example of multi-hop multi-stream video aggregation. White part indicates source symbols and shaded part indicates error protection symbols. Optimal: The amount of source symbols to achieve optimal visual quality. Available: The amount of source symbols received from the previous hop.

segment. We use $a_{j,l}$ to represent the number of source symbols of stream j that is allocated for segment l. If a segment l is assigned to stream j, we have $a_{j,l} > 0$. Otherwise, $a_{j,l} = 0$. The overall segment-to-stream assignment can be represented as a $J \times L$ matrix $\mathbf{A}$ with $[\mathbf{A}]_{j,l} = a_{j,l}$. Note that in order to reduce the communication overhead of representing the assignment pattern, we assume each segment can only be assigned to one stream. Therefore, there is only one non-zero element in each column of $\mathbf{A}$. The number of non-zero elements in j^{th} row of $\mathbf{A}$ is then the number of segments we determine to assign to stream j.

We define NZ() as the function that obtains the number of non-zero elements in a column or row vector. For PDMA-based multi-stream video aggregation strategy, we have $NZ(ac_l) = 1, \forall l \in \{1, 2, ...L\}$, where ac_l denote the l^{th} column vector of $\mathbf{A}$. Bandwidth constraint for multi-stream aggregation can be represented as $\sum_{j=1}^{J} L_j = L$, where $L_j = NZ(ar_j)$ and ar_j denotes the j^{th} row vector of $\mathbf{A}$. For unequal error protection, we apply stronger RS codes for more important data, i.e., $a_{j,l1} \leq a_{j,l2}$, if $l1 < l2$ and segments $l1$ and $l2$ are allocated to the same video stream.

To provide good video quality to all end users as well as fairness across users, we formulate the multi-hop multi-stream video aggregation problem for hop k as the following min-max optimization problem:

$$\min_{\mathbf{A}}(\max_{\{u\in\{1,2,...U\},j\in\{1,2,...J\}\}}(w_{u,j} \cdot ED_{u,j}(\mathbf{A}))) \tag{5.1}$$

subject to

$$\begin{cases} NZ(ac_l) = 1, \forall l \in 1, 2, ...L \\ NZ(ar_j) = L_j, \sum_{j=1}^{J} L_j = L \\ a_{j,l1} \leq a_{j,l2}, \ \ if \ a_{j,l1} > 0, \ a_{j,l2} > 0, \ and \ l1 < l2 \\ \sum_{l=1}^{L} a_{j,l} \leq Rp_j \end{cases}$$

Here, $ED_{u,j}$ is the expected distortion of stream j received by end user u, and Rp_j is the rate of source symbols correctly received from the pervious hop. $w_{u,j}$ is the user preference factor of user u for stream j. The larger $w_{u,j}$ indicates that user u desires higher visual quality of stream j. It is normalized for each user u, i.e. $\sum_{j=1}^{J} w_{u,j} = 1$.

In (5.1), the fourth constraint is the multi-hop constraint. It indicates that the source symbols available for assignment are the successfully FEC decoded source symbols received from the previous hop. For a compressed video stream, because of encoding dependency, some source symbols are dependent on previous part of source symbols in the same video stream to provide useful information to reduce the decoded video stream distortion. In a multi-hop environment, if the successfully FEC decoded source symbols received from previous hop cannot contribute to reduce visual distortion of the video stream, these source symbols should not be further aggregated. Therefore, Rp_j in (5.1) only represents the rate of source symbols that can contribute for reducing video stream distortion. For MPEG-4 FGS, source symbols are obtained bit-plane by bit-plane. If a source symbol is lost, the contribution of source symbols following this source symbol is negligible to reduce the video stream distortion. By exploiting UEP in the stair-case fashion [7, 50] to FGS layer, Rp_j, the rate of source symbols that can contribute for reducing video stream distortion, can be approximated to the rate of successfully FEC decoded source symbols.

5.1.1 Multi-hop Error Propagation

To solve the formulated problem (5.1), for each hop, we need to determine the PDMA-based error protection pattern. Since the target function of (5.1) is the

expected distortion of the streams that end users consume instead of the receiver of current hop, different kinds of communication hops may have different strategies to reduce the visual distortion of the streams that end users view. We can categorize communication hops to two different types and solve (5.1) with different approaches for each type of hops.

The first kind of hops is a communication link that directly transmits streams to an end user. We refer to them as EH hops. Since the video stream aggregated over this hop is directly consumed by end user, the error protection strategy and resource allocation for EH hops can be considered independent from other hops. The visual quality of streams that EH hops aggregate cannot affect the visual quality of streams that other hops aggregate in the system. To achieve the overall min-max visual quality of streams received by all end users, the best strategy of EH hops is just performing an optimized UEP and resource allocation algorithm based on its own transmission channel condition. The formulated multi-hop multi-stream problem (5.1) is then simplified to a single-hop multi-stream video aggregation problem as formulated in our previous study [11] in Chapter 2.

The second kind of hops is an intermediate communication link that does not directly transmit streams to an end user. We refer to them as IH hops. The video streams aggregated over this hop are further aggregated over other hops towards end users. For IH hops, we discuss two different cases for video stream aggregation: *overflow* and *underflow*. Given a fixed bandwidth, if the number of a video stream source symbols needed for the current hop's optimal FEC assignment is less than the number of source symbols correctly received from the previous hop, we refer to this condition as *overflow*. Otherwise, if the number of the source symbols needed for the current hop's optimal FEC assignment is larger, it is *underflow*. If *overflow*

happens, visual quality of this stream received by end users is not affected by the video aggregation of previous hop, rather determined by the optimal strategy of current hop. Based on this observation, if *overflow* happens in every intermediate hop in a system, the visual quality of delivered streams is only determined by the optimal strategy of EH hops. In this case, applying the single-hop multi-stream scheme [11] in EH hops can achieve the optimal min-max distortion for all video streams received by end users. However, when *underflow* happens, the EH hops do not have enough source symbols to be assigned to achieve the optimal expected distortion. The delivered visual quality is degraded from the optimal visual quality that EH hop can deliver if there are enough source symbols received. This hop-to-hop error propagation caused by *underflow* can significantly degrade the visual quality of streams received by end users in a multi-hop multi-stream environment.

5.1.2 Resource Allocation Strategy on Error-Free Channel

For the strategy of each kind of hops, we first consider error-free channel for simplicity, and then extend it to error-prone channel by considering error protection in addition to bandwidth allocation. We can reformulate problem (5.1) for error-free channels as follows:

$$\min_{\mathbf{A}}\Big(\max_{\{u\in\{1,2,...U\},j\in\{1,2,...J\}\}}(w_{u,j}\cdot D_j(R_{u,j}(\mathbf{A})))\Big) \tag{5.2}$$

subject to

$$\begin{cases} NZ(ac_l)=1,\forall l\in 1,2,...L \\ NZ(ar_j)=L_j,\sum_{j=1}^{J}L_j=L \\ a_{j,l}\in\{0,N\} \\ \sum_{l=1}^{L}a_{j,l}\le Rp_j \end{cases}$$

Here, $R_{u,j}$ is the source symbol rate of stream j received by end user u. We would like to achieve min-max optimization on $D_j(R_{u,j})$, the deterministic distortion instead of the expected distortion. This deterministic distortion is caused by the rate-shaping of video streams due to bandwidth limitation. Because there is no error protection symbols assigned for error-free channel, $\mathbf{A}$ has been simplified to only take two possible values, the length of segment N or zero.

For EH hops, this formulated problem can be easily solved by bi-section search on R-D curves of aggregated video streams [11]. We apply this approach to every EH hops, and denote the corresponding rate assignment solution as $\bar{R}_{u,j} = \sum_{l=1}^{L} a_{j,l}$ for stream j user u.

It is not trivial to solve (5.2) for IH hops. Since $R_{u,j}$ is the rate of video stream that end user receives, it does not have obvious relation to the rate we should assign for IH hops. This problem formulation needs further revising. For *underflow* cases, because the bandwidth bottleneck is at the IH hop to provide visual quality of stream j demanded by end users, we should maximize the delivered source symbol rate for stream j, i.e., minimize the delivered visual distortion for stream j at the IH hop. For *overflow* cases, the bottleneck is not at the IH hop, we should maximize the margin between EH demanded rate and IH provided rate to avoid the bottleneck. To facilitate our discussion, for an IH hop, we use Θ_j to denote the index set of end users that video streams j is aggregated to. For example, in the multi-hop multi-stream environment shown in Fig. 2.1, for IH hop C_{12}, $\Theta_1 = \{4, 5, 6, 7, 8, 9, 10\}$. Let DI_j be the comparison result of the rate provided by current IH hop and the rate demanded by EH hops for stream j. We define DI_j as:

$$DI_j = Comp(D_j(R_j(\mathbf{A})), \max_{\{u \in \Theta_j\}}(D_j(\bar{R}_{u,j})), \delta) \tag{5.3}$$

Here, $\bar{R}_{u,j}$ is the rate of video stream j that end user u demanded for EH hop to

achieve the min-max optimal distortion across received streams. To obtain $D_j(\bar{R}_{u,j})$, we solve (5.2) for the EH hop that aggregates streams to user u, assuming there are enough source symbols provided for each aggregated stream. R_j is the rate assigned in the current hop. We define a function Comp( ) to compare EH demanded rate and IH provided rate, and the output of Comp( ) is to indicate the objective of our strategies for *overflow* and *underflow* cases:

$$Comp(x, y, \delta) = \begin{cases} x - \delta y, & if\ x - \delta y \leq 0 \\ x, & if\ x - \delta y > 0 \end{cases} \tag{5.4}$$

$\delta \in [0, 1]$ is a relaxation parameter that can be set by system administrator of a communication system to adjust the usage scale of $D_j(\bar{R}_{u,j})$, i.e., the feedback distortion value to IH hops from EH hops. By defining DI_j, we can reformulate (5.2) as:

$$\min_{\mathbf{A}}(\max_{\{j\in\{1,2,...J\}\}}(\max_{\{u\in\Theta_j\}}(w_{u,j})\cdot DI_j)) \tag{5.5}$$

subject to

$$\begin{cases} NZ(ac_l) = 1, \forall l \in 1, 2, ...L \\ NZ(ar_j) = L_j, \sum_{j=1}^{J} L_j = L \\ a_{j,l} \in \{0, N\} \\ \sum_{l=1}^{L} a_{j,l} \leq Rp_j \end{cases}$$

We find that for $\delta = 1$, if every IH hop in a system seeks the optimal solution of (5.5) and EH hop seeks the optimal solution of (5.2), the system should achieve the optimal min-max distortion as defined in (5.2). This can be shown by contradiction as follows:

Assume every IH hop in a multi-hop multi-stream aggregation system achieves the optimal solution of (5.5), and every EH hop achieves the optimal solution of

(5.2). We have the overall min-max distortion as $D^{\diamond} = w_{u^{\diamond},j^{\diamond}} \cdot D_{j^{\diamond}}(R_{u^{\diamond},j^{\diamond}})$. Suppose there exists another optimal min-max distortion $D^{\star} = w_{u^{\star},j^{\star}} \cdot D_{j^{\star}}(R_{u^{\star},j^{\star}})$, and $D^{\star} < D^{\diamond}$, we can discuss the following three cases for contradictions:

Case 1: If there is no *underflow* in the EH hop which aggregates stream $j^{\diamond}$ to user $u^{\diamond}$, $D^{\diamond}$ is then the maximum distortion of video streams aggregated over this EH hop to user $u^{\diamond}$. $D^{\star} < D^{\diamond}$ is contradict to our assumption that every EH hop achieves the min-max distortion as defined in (5.2).

Case 2: If there is *underflow* in the EH hop which aggregates stream $j^{\diamond}$ to user $u^{\diamond}$, but not in the stream $j^{\diamond}$, we still have $D^{\diamond}$ as the maximum distortion of video streams aggregated over this EH hop. This results in the same contradiction as Case 1.

Case 3: If there is *underflow* of stream $j^{\diamond}$ for user $u^{\diamond}$ in EH hop, there must exist a hop k, such that

$$D^{\diamond} = w_{u^{\diamond},j^{\diamond}} \cdot D(R_{u^{\diamond},j^{\diamond}}) = \max_{\{j \in \{1,2,\ldots J\}\}} [\max_{\{u \in \Theta_j\}} (w_{u,j}) \cdot D_j^k(R_j^k)]. \tag{5.6}$$

Here, D_j^k is the distortion of stream j aggregated over hop k and R_j^k is the corresponding rate the receiver of hop k received. Based on the *underflow* assumption, we must have $D_{j^{\diamond}}^k(R_{j^{\diamond}}^k) < \max_{u \in \Theta_{j^{\diamond}}}(D_{j^{\diamond}}(\bar{R}_{u,j^{\diamond}}))$. Therefore, the output of function *Comp()* in (5.5) is $D_{j^{\diamond}}^k(R_{j^{\diamond}}^k)$. As the result, $D^{\diamond}$ is the min-max distortion achieved by solving (5.5) for the aggregation hop k. However, $D^{\star} < D^{\diamond}$ indicates the aggregation hop k can achieve lower distortion. $D^{\star} < D^{\diamond}$ is then contradict to our assumption that every IH hop in a multi-hop multi-stream aggregation system achieves the min-max distortion as defined in (5.5).

Based on the analytical study of Cases 1-3, $D^{\diamond}$ is the optimal min-max video stream distortion for the multi-hop multi-stream aggregation system.

5.1.3 Multi-Hop Multi-Stream Aggregation over Error-Prone Channel

Based on (5.5), by substituting deterministic distortion by expected distortion, and rate-based resource allocation by resource allocation that considers PDMA-based FEC, we now arrive at the problem formulation of multi-hop multi-stream video aggregation over packet erasure channel:

$$\min_{\mathbf{A}}(\max_{\{j\}}(\max_{\{u\in\Theta_j\}}(w_{u,j})\cdot Comp(ED_j(\mathbf{A}),\min_{\{u\in\Theta_j\}}(\bar{ED}_{u,j}),\delta))) \tag{5.7}$$

subject to

$$\begin{cases} NZ(ar_l)=1,\forall l\in 1,2,...L \\ NZ(ac_j)=L_j,\sum_{j=1}^{J}L_j=L \\ a_{j,l1}\le a_{j,l2},\ \ if\ a_{j,l1}>0,\ a_{j,l2}>0,\ and\ l1<l2 \\ \sum_{l=1}^{L}a_{j,l}\le R_j \end{cases}$$

Here, $\bar{ED}_{u,j}$ is the expected distortion of stream j for the end user u if there is no *underflow*. It can be obtained by assuming there are enough source symbols provided for optimal FEC assignment of each stream and using the single-hop multi-stream scheme described in Chapter 2 for every EH hops. $\bar{ED}_{u,j}$ is then feedback to IH hops. By setting $\bar{ED}_{u,j}$ to zero, (5.7) becomes equivalent to single-hop multi-stream problem formulation (5.1) for EH hops. Therefore, we can use (5.7) as a unified problem formulation for both IH hops and EH hops in multi-hop multi-stream aggregation.

5.2 Proposed Algorithm

5.2.1 An Iterative Search Algorithm

We now propose an iterative search algorithm to solve (5.7). This algorithm starts from an initial point which is close to optimal solution and iteratively moves towards the direction of reducing min-max expected distortion.

Step 1: Initialization

For a successful iterative searching algorithm, it is critical to start with an initial point that is close to optimal solution. This initial point for our formulated problem (5.7) is the segment assignment $\{L_j^{(0)}\}$ for PDMA-based error protection. We can obtain this initial point by using the bi-section search on R-D curves similar to bi-section search described in Section 3.2.3.

Step 2: Coarse search

In this step, we determine the searching direction towards optimal PDMA-based FEC segment assignment. For each segment assignment L_j with the available source symbol rate R_j, we can obtain an expected distortion for every stream j:

$$EDV_j = \max_{\{u\in\Theta_j\}}(w_{u,j}) \cdot Comp(ED_j(\mathbf{A}), \min_{\{u\in\Theta_j\}}(\bar{ED}_{u,j}), \delta). \tag{5.8}$$

Assuming that we have performed k iterations, we will find:

$$\begin{aligned} j_{max} &= \arg\max_j(EDV_j), \\ j_{min} &= \arg\min_j(EDV_j), \\ ED_{max}^{(k)} &= \max_{\{j\}}(EDV_j). \end{aligned} \tag{5.9}$$

We then take one segment from stream j_{min} and give one more segment to the

stream j_{max}:

$$\begin{aligned} L_{j_{max}}^{(k+1)} &= L_{j_{max}}^{(k)} + 1 \\ L_{j_{min}}^{(k+1)} &= L_{j_{min}}^{(k)} - 1 \\ L_{j}^{(k+1)} &= L_{j}^{(k)}, \forall j \neq j_{max}, j \neq j_{min} \end{aligned} \tag{5.10}$$

The corresponding expected distortion for the two streams that are involved in exchanging segments, namely, $EDV_{j_{max}}$ and $EDV_{j_{min}}$, are also updated. The above coarse search procedure is repeated until $EDV_{max}^{(k+1)} \geq EDV_{max}^{(k)}$.

Step 3: Fine search

In coarse search, we only examine the video streams with the minimum and maximum expected distortion, and change segment assignments to these two streams. In this step, we perform another round of finer iterative search that considers all of aggregated video streams to obtain the min-max expected distortion. Let $EDV_{max}^{(k-1)}$ be the maximum distortion in the $(k-1)^{th}$ iteration. In the k^{th} iteration, we perform $J-1$ trials by taking one segment from stream j, $j \neq j_{max}$, and assigning one more segment to stream j_{max}. In the j^{th} trial, if the updated maximum distortion is smaller than $EDV_{max}^{(k-1)}$, it implies that there exists a better solution than the result in the $(k-1)^{th}$ coarse search iteration. We will then use the segment assignment in trial j as a new start-point, and move to next iteration to perform coarse search again. Otherwise, with a total of $J-1$ trials of such examination, if the maximum distortions in all trials are not smaller than $EDV_{max}^{(k-1)}$, the finer searching is complete in the k^{th} iteration and $\{L_j^{(k-1)}\}$ is the optimal segment assignment.

Note that the optimal solution to (5.7) is not unique and there may exist several sets of solutions with the same min-max distortion. Using the proving-by-contradiction approach [11], we can show that there does not exist a better solution that can achieve smaller maximum distortion across streams than the one achieved

by our proposed algorithm.

5.2.2 Practical Implementation Issues

The key idea of multi-hop awareness is that the pervious hops a video stream is aggregated over are aware of the condition of the following hops that the same video stream is going to be aggregated over. This awareness is generally enabled by feedback. In our formulated problem (5.7), every hop needs to be aware of the end user optimal expected distortion $E\bar{D}_{u,j}$ to achieve the overall min-max expected distortion for multi-hop multi-stream system. However, in the practical use, there are obstacles to obtain $E\bar{D}_{u,j}$. First, some applications of multi-hop multi-stream aggregation, such as video conferencing, is real-time. To obtain $E\bar{D}_{u,j}$, we need to obtain the R-D information of video streams and channel condition of EH hops, and then feedback expected distortion information to IH hops. This feedback may create a delay too long to be acceptable for real-time applications. Second, the channel condition is varying, it results in the varying optimal expected distortion. Despite these obstacles, we can take advantage of the strong correlations between frames to alleviate these problems since the video stream R-D characteristics vary quite slowly from one frame to the next frame except at scene changes. In addition, the PDMA-based error protection scheme can further reduce the frame-by-frame variation of visual quality by exploring heterogeneity in R-D characteristics across multiple streams. Therefore, we can feedback the expected distortion $E\bar{D}_{u,j}$ of previous video frames from EH hops to IH hops, then use these feed back values in (5.7) to obtain optimal solution of current frame. Since most feedback channel is low rate, we average the optimal expected distortion over a certain number of frames (30 in our experiments), then feed it back to IH hops.

Table 5.1: The optimal expected distortion of video streams averaged over every 30 frames

Stream Index	1	2	3	4	5
PSNR(dB) averaged over frames 1-30	30.85	30.95	30.68	30.87	32.90
PSNR(dB) averaged over frames 31-60	31.11	31.05	30.78	31.13	32.98
PSNR(dB) averaged over frames 61-90	30.96	30.97	30.72	30.96	32.97

To show the feasibility of this method, we perform an experiment on the video conferencing system as shown in Fig. 2.1. We set the bandwidth of uplink for each user as 3 Mbps and the bandwidth of downlink as low as 2-3 Mbps to avoid *underflow* condition. The bandwidth of IH hops is 14.4 Mbps. Input video streams is set to be the same as experiments described in Section 3.3.2. With the packet loss rate shown in Table 5.2 and user preference shown in Table 5.3, the optimal expected distortion of video stream 2-6 that user 1 received and averaged over every 30 frames is shown in Table 5.1. We can see that the expected distortion remains relatively stable with less than 0.27 dB difference in terms of PSNR. Other streams received by end users demonstrate similar characteristics.

In the practical implementation of the proposed algorithm, δ in (5.7) is adjusted to make our proposed scheme more adaptive to the varying channel condition and can be set by a system administrator based on the channel condition of a communication system. If there is no feedback channel bandwidth available or the channel condition is varying too fast to make the feedback value usable, the system administrator can set δ to be 0, so that the algorithm will be the same as performing

FEC and resource allocation with only local information of each hop. If δ is set to be 1, the feedback expected distortion will be fully used. When channel condition is varying relatively slower compared to the period between two feedback time, we should set relatively higher δ with $\delta \in (0, 1)$.

5.3 Experimental Results

In this section, we evaluate the effectiveness of our proposed multi-hop multi-stream video aggregation scheme (*MHMS*) by comparing it to an alternative single-hop multi-stream video aggregation scheme (*SHMS*) that we previously proposed in Chapter 2 [11, 14]. Since both video aggregation schemes use the same building blocks on source coding, channel coding and packetization, our comparison can evaluate the performance gain of multi-hop awareness in terms of visual quality of the delivered video streams.

Our system set-up is a distributed video conferencing system as shown in Fig. 2.1, and we assume that the channel for each hop is a packet erasure channel. Due to the spatial and temporal prediction in video compression, burst loss of packets of a compressed video stream can result in severe degradation of video stream visual quality. Packets interleaving is a common used technique to prevent burst packets loss. In terms of the delivered video quality, transmitting interleaved packets over a memory channel should be equivalent to transmitting packets in their original order as shown in Fig. 2.3 over a memoryless packet erasure channel. Therefore, without loss of generality, we can consider that packet loss is memoryless in our system. The users' input streams are the same as experiments described in Section 3.3.2. Each of these ten streams is encoded into 30 frames per GOP and each GOP is leading by one I frame followed by 29 P frames. The base-layer of

each stream is encoded with quantization parameter $Q = 30$. There are 8 bits per symbol and the encoded bit-stream packet size is 128 bytes.

The same frame interleaving technique as described in Section 3.3.1 is used in our simulation to avoid tremendous data rate fluctuation from frame to frame when multiple streams are merged together. Because of the un-symmetric data stream volume of the uplink and downlink from user, the bandwidth of uplink is usually much smaller than the downlink bandwidth. In our experiments, the bandwidth of uplink for each user is 3 Mbps and the bandwidth of downlink for each user is in the range of 6-9 Mbps. The bandwidth of IH hops is 14.4 Mbps. One user receives 90 frames of video streams from all other users. We would like to vary the user preference for different incoming video streams, so the user preference value in our experiments are set in eight different levels from 5 to 40 with step size 5. They are shown in Table 5.3.

We would like to vary the packet loss rate for different hops and test the packet loss rates which are common in various types of communication networks, i.e., less than 40%. Therefore, without losing the generality, we generate random numbers which are uniformly distributed on the set $\{0.1, 0.2, 0.3, 0.4\}$, and then use these random numbers as the packet loss rates of hops. The generated packet loss rates are listed in Table 5.2. We collect simulation results in 3 seconds of video aggregation, i.e. 90 video frames with 30 frames per second, and repeat multiple runs. The optimal expected distortion of EH hops is feeded back to intermediate IH hops every 1 second. Our experimental results are shown in Fig. 5.2 and Fig. 5.3. The frame-by-frame average PSNR across 9 received streams for user 3 are illustrated in Fig. 5.2. The result is averaged over repeated 100 test runs. We can see from Fig. 5.2 that *MHMS* can provide higher video quality than *SHMS*. The performance

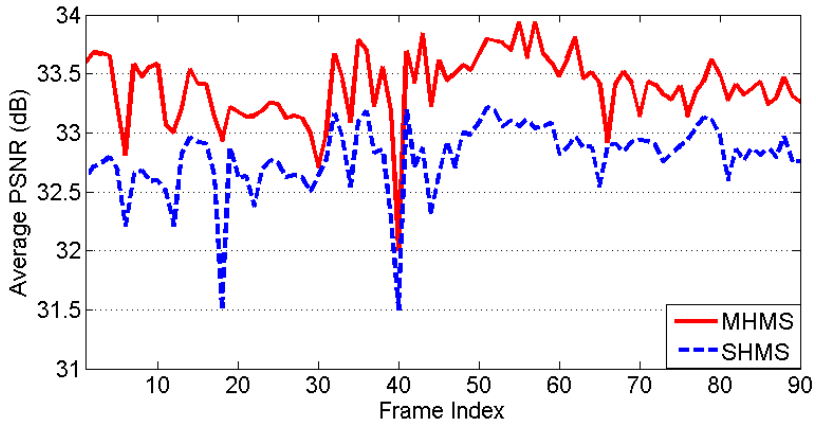


Figure 5.2: Performance comparison of MHMS vs SHMS schemes: Frame-by-frame PSNR averaged across 9 received streams for user 3 in Fig. 2.1. MHMS: FEC and resource allocation with multi-hop awareness ; SHMS: FEC and resource allocation without multi-hop awareness.

gain in terms of PSNR is up to 1.42 dB. The averaged PSNR across the frames is improved from 32.67 dB to 33.48 dB with 0.81 dB gain. We observe the similar performance gain in other user's received video streams as well, and the gain is up to more than 1 dB. To summarize the results, Fig. 5.3 shows the PSNR across 90 video frames averaged over 10 users with 9 received streams per user, and repeated 100 test runs. We can see that *MHMS* outperforms *SHMS* with PSNR gain up to 0.8 dB. The average PSNR gain is 0.61 dB.

In the previous experiment, δ in (5.7) is set to be 1 to take the full usage of feedback expected distortion. Our next experiment aims at demonstrating the effect of setting up different δ for a system. The video stream and channel condition set-up is the same as in the previous experiment, and Fig. 5.4 shows *MHMS* gain compared to *SHMS* with different δ in terms of the visual quality of user 3 received

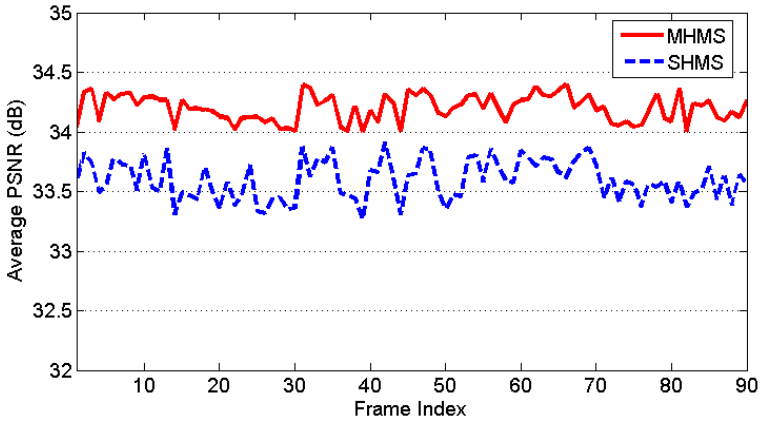


Figure 5.3: Performance comparison of MHMS vs SHMS schemes: Frame-by-frame PSNR averaged across 10 users with 9 received streams per user. MHMS: FEC and resource allocation with multi-hop awareness ; SHMS: FEC and resource allocation without multi-hop awareness.

video streams. We can see from Fig. 5.4 that, for a system with relatively slow-varying channel condition, the smaller δ will result in less multi-hop awareness gain. When δ is decreased below certain value (0.8 in Fig. 5.4), the feedback expected distortion scaled by δ is smaller than the distortion IH hop can provide and this results in *underflow*. When every stream in IH hop is *underflow*, *MHMS* problem formulated in (5.7) is equivalent to *SHMS* problem formulated in Section 3.1, so that *MHMS* and *SHMS* have the same performance. To set a proper δ for a multi-hop multi-stream aggregation system, a system administrator can perform a calibration procedure before the system enters the normal operation mode, i.e., setting different δ for a trial video conferencing session to find out a proper δ that could achieve best delivered video quality.

Table 5.2: The channel condition of communication hops shown in Fig. 2.1

Hop	U_1	U_2	U_3	U_4	U_5	U_6	U_7	U_8
Packet loss rate	0.1	0.3	0.2	0.1	0.2	0.2	0.1	0.1
Bandwidth (Mbps)	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
Hop	U_9	U_10	V_1	V_2	V_3	V_4	V_5	V_6
Packet loss rate	0.1	0.3	0.4	0.1	0.2	0.3	0.4	0.2
Bandwidth (Mbps)	6.9	6.2	6.0	7.8	3.0	3.0	8.4	9.0
Hop	V_7	V_8	V_9	V_10	C_{12}	C_{21}	C_{23}	C_{32}
Packet loss rate	0.3	0.2	0.4	0.1	0.2	0.4	0.3	0.1
Bandwidth (Mbps)	6.6	8.4	8.1	8.7	14.4	14.4	14.4	14.4

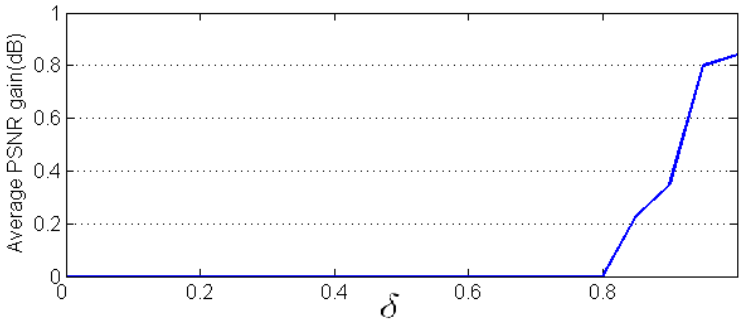


Figure 5.4: Performance comparison with varying δ.

Table 5.3: The user preference for incoming stream from user US_x to receiving node UN_x . The video conferencing topology is shown in Fig. 2.1

Preference	US_1	US_2	US_3	US_4	US_5	US_6	US_7	US_8	US_9	US_{10}
UN_1	0	5	5	5	5	5	5	10	30	30
UN_2	5	0	5	30	30	5	10	5	5	5
UN_3	5	5	0	5	5	5	30	30	10	5
UN_4	10	15	5	0	10	15	5	5	30	5
UN_5	5	10	20	5	0	10	10	10	5	25
UN_6	20	30	5	15	5	0	5	5	5	10
UN_7	5	5	15	5	20	5	0	15	20	10
UN_8	5	10	20	5	10	5	30	0	10	5
UN_9	20	5	5	40	5	10	5	5	0	5
UN_{10}	5	10	5	5	30	30	5	5	5	0

5.4 Chapter Summary

This chapter investigates a multi-stream video aggregation scheme with multi-hop awareness over packet erasure channels. This video aggregation scheme explores multi-hop awareness, multi-stream heterogeneity and performs optimization in FEC and resource allocation to minimize the maximum distortion across all video streams delivered to all end users. Comparing to the multi-stream video aggregation scheme without multi-hop awareness, our simulation shows that the proposed scheme has significant gain in terms of the perceptual quality of delivered video streams. It is a promising shceme to support multi-stream video aggregation over packet erasure channel in a multi-hop environment.

Chapter 6

Classification-Based Error Concealment

Due to various kinds of distortion and failures, part of a compressed image or video can be damaged or lost during transmission. When the transport layer mechanism, such as FEC approaches discussed in the previous chapters, cannot provide sufficient error protection on the payload stream, the unrecovered transmission errors may lead to visual distortions at the decoder. The widely used block-based visual coding systems have prompted a need of block-based error concealment on the decoder side. A number of concealment approaches have been proposed in recent years [67, 69, 60, 68, 76, 35, 78, 3]. The smoothness and continuity properties in spatial or frequency domain, the repeating patterns, and other properties of visual data have been exploited to recover corrupted blocks from the survived surroundings. Through a benchmarking effort on the existing error concealment approaches, we have observed that different approaches are suitable for different image characteristics of a corrupted block and its surroundings, and none of the existing approaches is an all-time champion. This motivates us to explore a classification-based concealment approach that can combine the better performance of two state-of-the-art approaches in the literature. The classification-based approach also helps us achieve

a better tradeoff between the concealment quality and the computation complexity on the receiver side. This is because some state-of-the-art approaches have rather high computation demand, and classification allows the computation power to be spent more strategically by performing expensive computations only when they are likely to offer a substantial gain in concealment quality.

The classification in the proposed new framework of error concealment can be done either on the receiver side or on the sender side. The receiver-side classification uses the survived surrounding pixels to determine which candidate concealment approach would give better concealment quality for each corrupted block. As shall be seen in this chapter, the proposed receiver-side classification approach does not require side information and the overall concealment quality can outperform each candidate alone. To provide more proactive protection and further exploit the knowledge from the original, uncorrupted image, a few recent works in the literature [73, 74, 9, 42] have jointly considered the design of sender and receiver systems to facilitate error concealment. We explore this sender-driven perspective for our classification-based concealment framework by obtaining a small amount of classification data on the sender side. As the classification results need to be delivered as side information from the sender to the receiver, we examine and compare two approaches for delivering the side information, namely, by attaching as part of the file header and by embedding in the image signal.

The chapter is organized as follows. Section 6.1 provides a brief description of the evaluated algorithms and presents benchmarking results on a collection of natural and artificial images. Since the performance on various images shows the advantages and disadvantages of different error concealment techniques, a classification scheme on the receiver side is proposed in Section 6.2 to take advantages

of the sweet spots of existing techniques. The sender-side classification-based error concealment is proposed in Section 6.3 to further improve the concealment quality by supplying the ground-truth of concealment technique selection to a receiver. We compare the concealment performance, computation complexity, and bandwidth usage of the three proposed schemes as well as their suitable application scenarios in Section 6.4, and conclude this chapter in Section 6.5.

6.1 Background and Motivation

6.1.1 Prior Work

Early explorations on spatial domain image concealment were reviewed in [67]. Among them, the multi-directional interpolation (MDI) approach performs pixel-domain interpolation along eight possible edge directions and considers the cases of both single edge and multiple edges [69]; the projection-onto-convex-sets (POCS) approach constrains the feasible solution set based on such prior information as smoothness and neighborhood consistency [60]; and the maximally smooth recovery (MSR) method makes use of the smoothness property of visual signals and formulates the concealment as a constrained energy minimization problem [68].

Three recent works in [76, 35, 78] have demonstrated performance improvement on classic images such as "Lena" or "Barbara" over the earlier approaches. The geometric-structure-based (GSB) error concealment by Zeng et al. [76] is a directional interpolation scheme, which uses the local geometric information extracted from the surroundings. Two layers of pixels surrounding a corrupted block are converted to a binary pattern to reveal the local geometric structure and to classify the block as flat or non-flat. For flat blocks, the projective interpolation technique

of [29] is applied. For non-flat blocks, the edges inside the lost block are estimated by pairing significant transition points from the aforementioned binary pattern, and the lost pixels are recovered by bilinear interpolation along the edge directions.

The orientation adaptive sequential interpolation (OASI) scheme by Li et al. [35] employs a linear regression model. It first estimates the local characteristics from a neighborhood of about four layers of uncorrupted pixels, and then uses the model parameters obtained to estimate each missing pixel from its surrounding pixels. More specifically, the interpolation can be characterized by $S = \sum_{k=1}^{N} \alpha_k s_k$, where S is an estimate of the missing pixel and s_k's are N neighboring pixels. The interpolation coefficients α_k form a vector α, which can be determined using the classical least-square method from an M-pixel neighborhood M_n with $M > N$, i.e., $\alpha = (C^T C)^{-1} C y$. Here, y is an $M \times 1$ vector representing M pixels in the training area M_n; C is an $M \times N$ matrix, and each of its M rows consists of N neighbors around the corresponding pixel in y. When $C^T C$ is singular, α_k is set to $1/N$.

The long range correlation scheme (LRC) by Zhang et al. [78] exploits the repeating patterns in an image. It extracts a ring window surrounding the corrupted area, searches for an area in the image that best matches the pattern of this ring in a mean-squared error sense, and replaces the corrupted area with the pattern inside the best matching ring. Long range correlation is also exploited in the recent image inpainting work by Bertalmio et al. [3], where the basic texture synthesis procedure for concealing the lost area is similar to the LRC concealment algorithm. By simultaneously filling in the structure and texture information of missing areas, the inpainting technique demonstrates excellent subjective quality when the missing area is relatively small compared with the size of the whole image. It is worth noticing that the image inpainting technique focuses more on the overall subjective

quality and is not designed to optimize an objective error measure of the concealment quality (such as MSE or PSNR) on many small blocks.

6.1.2 Performance Benchmarking

If an image is compressed by a block-based codec and transmitted over an error-prone channel, the error impairments are likely to be in the block domain. We focus on isolated block concealment in this work because block-based codecs are dominant for image or video transmission and the interleaving techniques can be employed in packetization to prevent the consecutive block loss [74, 37, 6]. Since different error concealment techniques employ quite different "philosophies," it was not conclusive from the literature which one is the best. We attempt to address this issue through a benchmarking effort, which also sheds light on the design direction of a new concealment framework that can outperform the existing approaches.

We use a collection of 15 8-bit gray-scaled images with different characteristics to evaluate the performance of the six approaches reviewed above, namely, MDI, POCS, MSR, GSB, OASI, and LRC. The names and the corresponding references for these approaches are listed in Table 6.1. The collection of 15 images is shown in Fig. 6.11. They can be divided into roughly four categories according to the visual content, namely, portraits, artificial images, natural scenery images, and rich texture images. We test the concealment on a typical loss pattern as shown in Fig. 6.1, where a total of 25% blocks are lost in a checkerboard fashion and the block size is 8×8. This damage pattern is used throughout all the following experiments if not specified otherwise. We examine the quality of recovered images in terms of PSNR and the computation complexity in terms of the concealment speed, and summarize the results in Table 6.2 and Table 6.3, respectively. All algorithms have

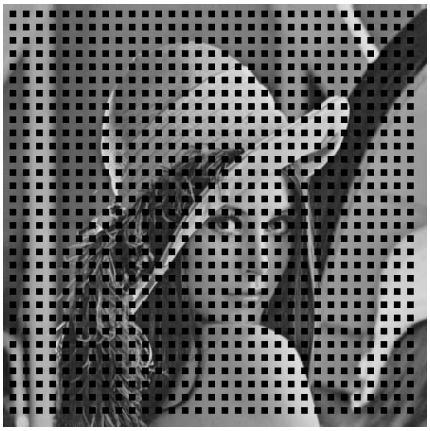


Figure 6.1: A checkerboard pattern with 25% block loss used in the concealment experiments.

been implemented in C/C++ with a moderate amount of optimization and the same speed-up settings, and tested on a 1.20 GHz Pentium-4 PC.

We can see from Table 6.2 that among the three recent techniques reviewed earlier, the LRC approach does not outperform the GSB and OASI approaches on most images. One reason is that the checkerboard error pattern leaves a very limited number of the candidate matching windows that do not suffer from the loss. The LRC approach does not perform well on most natural scenery images, either, since there is few repeating pattern. On the other hand, the GSB and OASI approaches significantly outperform other approaches on these benchmark images, although neither of the two gives the best performance for all images. The lack of all-time champion suggests that the image characteristics vary significantly from one to another, so a single algorithm based on an assumption about one aspect of the characteristics is not suitable for all images. This motivates us to go one step further and assemble a recovered image in which each concealed block is the better

Table 6.1: The names and the references for the benchmarked approaches

Acronym	Name	Reference
MDI	Multi-Directional Interpolation	[69]
POCS	Projection-Onto-Convex-Sets	[60]
MSR	Maximally Smooth Recovery	[68]
GSB	Geometric-Structure-Based	[76]
OASI	Orientation Adaptive Sequential Interpolation	[35]
LRC	Long Range Correlation	[78]

one selected between the GSB and OASI concealment results. As shown in the last column ("Better-2") of Table 6.2, this assembled image gives a much higher overall concealment quality than using GSB or OASI alone.

In terms of computation complexity measured in concealment speed, Table 6.3 shows that MSR and GSB are the fastest. MDI and OASI are about an order of magnitude slower, and LRC and POCS are by far the slowest algorithms. Jointly considering the concealment quality and speed, we see that although GSB and OASI both have high performance on concealment quality, OASI has relatively high computation complexity. If we could choose the OASI method to conceal corrupted blocks only when it provides significant performance gain, we would achieve both higher concealment quality and relatively lower computation complexity. This motivates us to research on an adaptive scheme for selecting error concealment methods to combine the advantages of these two good performing schemes.

Table 6.2: Comparison of algorithms in concealment quality PSNR (dB). The scheme achieving better performance between GSB and OASI is highlighted in bold italic font. The Better-2 column lists the concealment quality of recovered images in which each concealed block is the better one selected between GSB and OASI. Type A: Natural; Type B: Portrait; Type C: Artificial; Type D: Texture.

Type	Name	MDI	POCS	MSR	LRC	GSB	OASI	Better-2
A	Bassharbor	29.47	28.12	28.83	27.84	***30.69***	30.37	31.46
	Blueflower	27.88	27.55	27.09	26.77	29.68	***29.85***	31.04
	House	28.78	26.08	27.00	26.86	29.47	***30.00***	30.98
	NewYork	24.25	21.00	23.66	22.80	24.13	***24.52***	25.29
	Operahouse	30.91	28.88	28.53	29.08	30.88	***31.30***	32.38
	Papermachine	29.77	28.46	25.80	31.78	***33.85***	33.75	36.12
	Watch	31.40	29.59	29.41	31.35	33.77	***33.99***	35.52
B	Lena	32.28	29.49	29.20	30.64	34.43	***35.12***	36.08
	Barbara	27.41	23.35	27.14	29.78	29.26	***30.79***	31.80
	Kid	31.86	29.62	29.57	30.21	***33.47***	33.45	34.98
	Man	27.59	25.41	26.07	25.60	28.77	***29.13***	30.12
C	Circletrain	41.62	34.16	32.11	46.51	***48.33***	34.90	48.33
	Tulip	29.74	28.05	26.71	27.61	33.22	***33.47***	35.13
	Waterfall	27.92	26.36	26.52	26.18	28.79	***29.12***	30.20
D	Bear	30.05	29.55	27.99	27.82	32.33	***33.30***	34.38

Table 6.3: Comparison of algorithms in speed (seconds) for concealing the "Lena" image. All algorithms are tested on a 1.20 GHz Pentium-4 PC.

	MDI	POCS	MSR	LRC	GSB	OASI
Lena	3.03	219.58	0.59	98.45	0.56	7.12

6.1.3 Classification Based Concealment

For a receiver to pick the better one between the two state-of-the-art techniques correctly is a nontrivial task. This is because a receiver does not have the original undamaged image to compare with and determine which scheme gives better performance. Available to a concealment system are only the survived pixels that surround each corrupted block. If we could establish the connection between the image characteristics of the survived surrounding pixels and the correct selection between GSB and OASI using a training set, we could make a smart decision on which scheme to choose for a new damaged image.

To help exploring a rule in classifying the survived surrounding pixels, we take a close look at the "Better-2" test from Table 6.2. For each block, we quantify the error concealment performance of GSB and OASI by

$$\begin{aligned} P1 &= \textstyle\sum_{i=1}^{K} |C1_i - O_i|, \\ P2 &= \textstyle\sum_{i=1}^{K} |C2_i - O_i|, \end{aligned} \tag{6.1}$$

where K is the number of pixels in the block and is 64 in our case; O_i is the original value of the i^{th} pixel in the block; $C1_i$ and $C2_i$ are the corresponding recovered pixel values by GSB and OASI, respectively. We visualize in Fig. 6.2 the scheme selection for each lost block of the "Lena" image. The gray blocks indicate that GSB

Figure 6.2: Illustration of better performing concealment schemes between GSB and OASI on the "Lena" image: OASI performs better (white blocks); GSB performs better (black blocks); GSB and OASI do not have significant performance difference (gray blocks).

and OASI do not have significant performance difference (i.e., $|P1 - P2| < 96$); the white blocks indicate $P2$ is much smaller for the corresponding blocks; and the black blocks indicate that $P1$ is much smaller. From Fig. 6.2, we do not observe any obvious trend in determining where GSB and OASI would perform better: the black blocks appear in both edges and some texture areas, and so do the white blocks.

We further explore if one could deduce some simple rules from the spatial characteristics of survived pixels surrounding the lost blocks. We define a smoothness feature from four layers of survived surrounding pixels as follows. First, we group the pixels into a total of 48 segments, and each segment has 2×2 pixels, as shown in Fig. 6.3(a). For each segment, we generate a binary value characterizing smooth-

ness: if the range of the pixel intensity in the segment exceeds a pre-determined threshold of 15, we use "1" to indicate it as a non-flat segment; otherwise, we use "0." Next, the binary values from different segments are scanned according to the order in Fig. 6.3(b) to form a feature vector, which is a binary sequence. We count the total number of ones in the feature vector (i.e., the number of non-flat segments) for each of the 15 images used in our benchmark test. For each possible count of non-flat segments, we also compute the ratio of the number of blocks where OASI performs better versus those where GSB performs better. The relation is visualized in Fig. 6.4, where we can see a general trend that GSB is likely to perform better on smooth blocks, and OASI tends to be better for texture blocks. But, the curve is not monotonic and the ratios do not deviate much from one, suggesting that we cannot reliably determine the better performing concealment scheme just based on the non-flat segment count of the surviving surroundings.

The difficulty for a receiver in arriving at a simple rule to determine the better performing scheme can be tackled in two ways. If the decision is to be made solely on the receiver side, there is a need of employing advanced classification tools to group all possible surrounding pixel patterns into two classes, one class favoring the use of OASI for concealment, and the other class favoring GSB. Alternatively, we can avoid the difficult task of receiver-side classification by determining the classification information on the sender side where the uncorrupted image is available for providing ground-truth, and by sending such extra information to the receiver through attachment or data embedding techniques. In the next two sections, we will present the details of the proposed receiver-side and sender-side schemes, respectively. While we use OASI and GSB as building blocks to investigate our proposed framework of classification-based concealment, the new framework is general so that it can be eas-

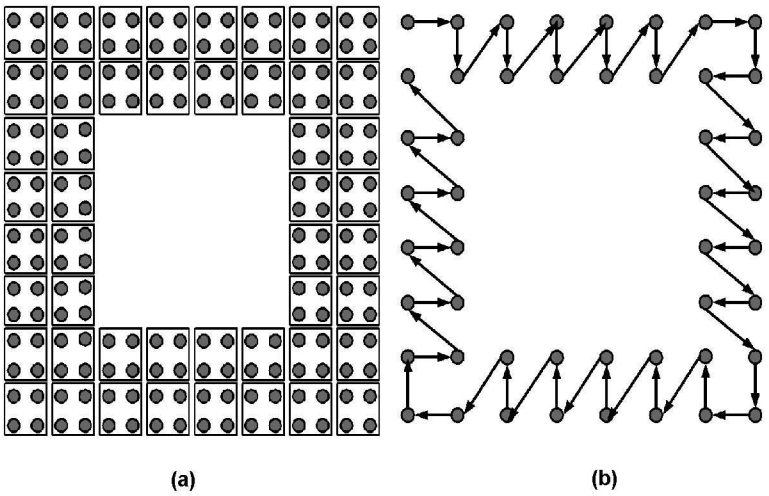


Figure 6.3: Feature extraction from survived surrounding pixels: (a) grouping of survived pixels into small 2×2 segments. (b) scanning order for constructing a feature vector.

ily extended to incorporate other appropriate concealment schemes and perceptual criteria.

6.2 Receiver-Side Adaptive Block Concealment Using SVM Classification

6.2.1 Classification Based on Support Vector Machine (SVM)

We formulate a receiver's choice of concealment scheme for each block as a supervised classification problem. Each error concealment method is considered as a class, and a feature vector is extracted from the pixels that surround an image block. In the training stage, we collect a number of feature vectors from training images, and label every feature vector x_i with a ground-truth class corresponding to

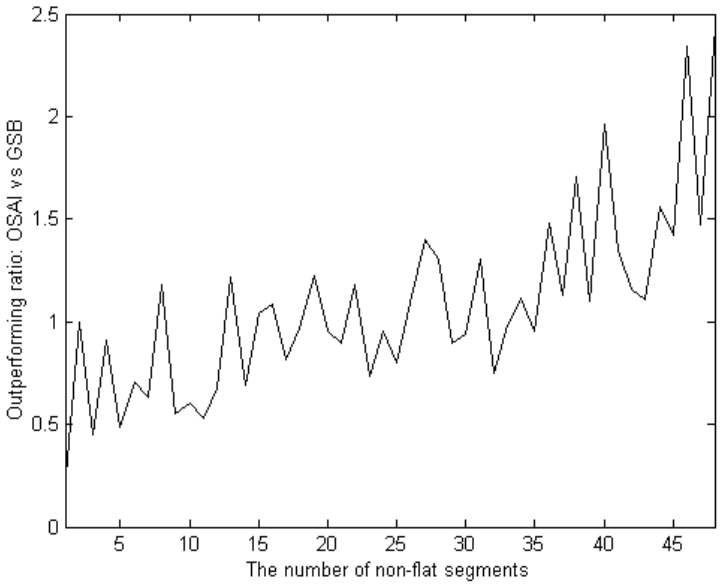


Figure 6.4: Examining the feasibility of a simple smoothness measure for distinguishing the better performing scheme: X-axis represents the number of non-flat segments in survived surroundings; and Y-axis represents the ratio of the block counts where OASI performs better to those where GSB is better.

the best concealment method for the associated block. We train the classifier using these feature-class pairs.

We adopt support vector machine (SVM) classifiers, as they often exhibit good generalization performance [19, 26, 40] with theoretical insights of structural risk minimization [4, 65]. The design of an SVM classifier can be boiled down to a convex quadratic programming problem with global optimal solutions in training. For our two-class pattern classification problem that decides between the GSB and OASI concealment approaches, two kernel functions have been used to search for the optimal classification solution, namely, a linear kernel function and a radial kernel function.

Linear SVM The linear SVM determines a linear discriminant function (a hyperplane) that gives the maximum separation margin between the two classes of training data [4]. The optimization problem can be formulated as

$$\text{minimize} \qquad f(\mathbf{w}, b) = \|\mathbf{w}\|^2, \tag{6.2}$$

$$\text{subject to} \qquad y_i(\mathbf{x}_i^T\mathbf{w} + b) - 1 \geq 0, \tag{6.3}$$

where $\mathbf{x}_i$ is the i^{th} training feature vector, and $y_i \in \{-1, 1\}$ represents the corresponding class label. The separating hyperplane is parameterized by a vector $\mathbf{w}$ and a scalar b, where $\mathbf{w}$ is the norm of the separating hyperplane. The Lagrangian-multiplier formulation for this constrained optimization problem is

$$L_p = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{l} \alpha_i y_i(\mathbf{x}_i^T\mathbf{w} + b) + \sum_{i=1}^{l} \alpha_i, \tag{6.4}$$

where $\{\alpha_i\}$ is a set of Lagrangian multipliers. Now, the problem is reduced to minimizing L_p with respect to w and b under the restrictions of (1) the derivatives of L_p

with respect to all α_i's vanish, and (2) $\alpha_i \geq 0$. For this convex quadratic programming problem, it is well established that the solution can be obtained through the Karush-Kuhn-Tucker (KKT) conditions, or through an easier *dual* problem [4].

When the training data of the two classes are linearly separable, the linear-kernel SVM approach gives a classifier in the form of a hyperplane separating the two classes of training data with the largest margin. If the training data are not linearly separable, a positive slack variable ξ_i ($\xi_i \geq 0$) can be introduced to alleviate the sensitivity of noisy training patterns [57]:

$$y_i(\mathbf{x}_i^T\mathbf{w} + b) - 1 + \xi_i \geq 0, \tag{6.5}$$

and

$$L_p = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i - \sum_{i=1}^{l}\alpha_i[y_i(\mathbf{x}_i^T\mathbf{w} + b) - 1 + \xi_i] - \sum_{i=1}^{l}u_i\xi_i, \tag{6.6}$$

where C is a parameter adjusting the relative penalty given to the classification errors on the training data.

To use a trained classifier to classify a new test sample $\mathbf{z}$, we evaluate the sign of the following function

$$f(\mathbf{z}) = \mathbf{w}^T\mathbf{z} + b = \sum_{i=1}^{N_s}\alpha_i y_i \mathbf{x}_i^T\mathbf{z} + b. \tag{6.7}$$

Here, $\mathbf{w}$ is explicitly determined by a set of N_s *support vectors*, which are such training vectors that lie closest to the hyperplane separating the two classes [4]. The sign reflects on which side of the decision boundary that $\mathbf{z}$ lies and thus determines the classification result.

Handling nonlinearity The feature vector as an input to a classifier for the concealment problem can be the pixel pattern surrounding a lost block, or some

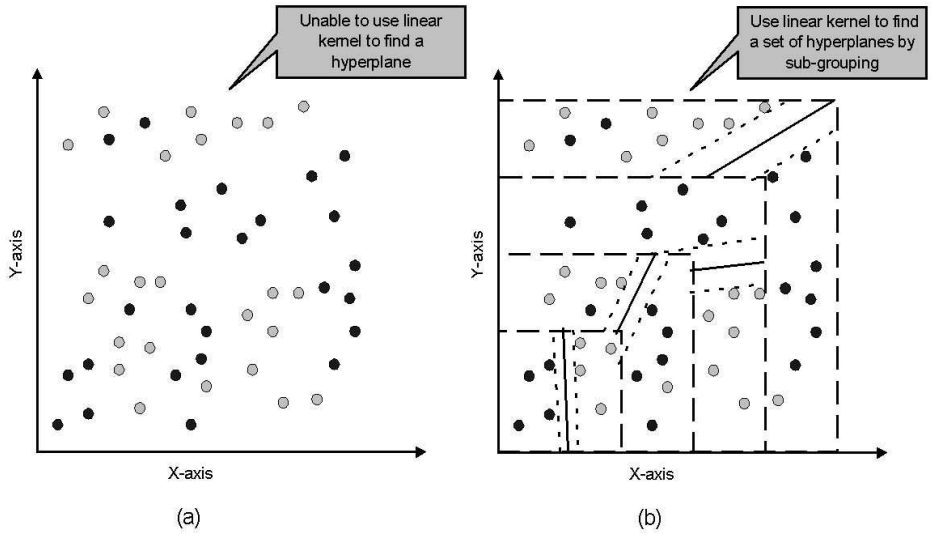


Figure 6.5: Handling the nonlinearity by a divide-and-conquer technique that trains a set of classifiers, one for each subset of the feature space.

statistics generated from the pattern (such as the binary feature vector defined in Section 6.1). The dimension of such feature vectors is rather high. Furthermore, the training features for each class may have complicated distributions, and in general are far from separable by a linear discrimination function in the original vector space. The non-separability by a linear discrimination function can be handled in two ways. One is extending the linear SVM with the kernel technique, and the other is dividing the vector space into groups and finding one classifier for each group.

Nonlinear classification functions [4] can be built by replacing the dot-product term $<\mathbf{x}_i, \mathbf{x}_j> = \mathbf{x}_i^T \mathbf{x}_j$ in the linear-kernel SVM by an appropriate kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$. This is equivalent to transforming feature vectors to a higher dimensional space H through a mapping $\Phi : R^d \rightarrow H$, and then finding a linear SVM classifier in this new space with $K(\mathbf{x}_i, \mathbf{x}_j) = <\Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j)>$. The radial basis kernel function

in the form of

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma^2} \tag{6.8}$$

is commonly used for its good generalization capabilities, especially when very limited information is available about the data distribution and separability for all classes. σ is the width of the radial basis. It affects the classification performance substantially, and will be addressed later in the section.

An alternative way to dealing with the nonlinearity is to use a divide-and-conquer technique. The idea is illustrated by the two-dimensional example shown in Fig. 6.5, where the two classes of data represented in Fig. 6.5(a) are not linearly separable. However, if we divide the space into four stripes as shown by the dashed lines in Fig. 6.5(b), the data within each stripe become more separable by a linear function. The subdivision of the feature space naturally accommodates the nonlinearity in the class boundary, yet the training process is comprised of training a set of relatively simple linear SVMs. Subdividing the feature space into non-overlapped subsets can be done through dividing the dynamic range of some feature elements, or according to the norm of the feature vector. The latter reflects the overall smoothness of the surrounding pattern for the feature vector defined in Section 6.1, as the L_1 norm of the vector gives the total number of non-flat 2×2 segments over the 48 pixel segments surrounding a lost block. Recalling the trend seen in Fig. 6.4 on the classes as a function of the overall smoothness, the subdivision allows us to naturally adapt to the changing cluster characteristics.

The nonlinearity in the classification can also be handled using a combination of the above two approaches. This hybrid approach divides the feature space into subsets and provides a nonlinear SVM (such as the radial kernel function) for each subset. It offers a great amount of flexibility, allowing the subsets to use dif-

ferent kernel parameters (such as σ in the radial basis function), or even different kernels. The nonlinear SVM obtained for each subset of feature space can have a much smaller number of support vectors, hence be considerably simpler than a nonlinear SVM trained for the entire space. As such, the hybrid approach has a low computation complexity in both the training and test phases.

Determining kernel parameters In practice, the relation between the classification accuracy on the training set and on the test set relies highly on the generalization capability of the classifier. In SVMs, there are several important parameters affecting the generalization capability, such as C in (6.6) and σ in (6.8). Choosing SVM kernel parameters can be viewed as a validation process, and evaluating the performance of the trained model on a validation set is a general approach to select kernel parameters. Based on this approach, we propose the following pre-processing procedure for choosing the kernel parameters.

Step-1 Dividing the training samples into two subsets, $\mathcal{A}$ and $\mathcal{B}$: in each iteration below, we use set $\mathcal{A}$ for training, and set $\mathcal{B}$ for testing.

Step-2 Choosing kernel parameters and constructing a new training set $\mathcal{R}$: we adjust kernel parameters $\sigma^{(1)}$ and $C^{(1)}$ so that the sum of training errors on $\mathcal{A}$ and test errors on $\mathcal{B}$ is minimized. More generally, we may employ an objective function using a weighted sum of the two types of errors, and low error rate on the test set is often desirable to ensure a good generalization capability of the classifier. Since SVM is known to generalize well and does not usually suffer from as much overfitting problem as the conventional classifiers, we choose to minimize the sum of errors (i.e. with equal weights) for simplicity. A new training set $\mathcal{R}$ is then generated consisting of the support vectors from set $\mathcal{A}$ and the successfully classified samples from set $\mathcal{B}$.

Step-3 Switching subsets: we switch set $\mathcal{A}$ with set $\mathcal{B}$ and repeat *Step 2*. We record the kernel parameters as $\sigma^{(2)}$ and $C^{(2)}$, and denote the new training set as $\mathcal{S}$. The union of set $\mathcal{R}$ and set $\mathcal{S}$ becomes the final training set $\mathcal{T}$.

Step-4 Determining kernel parameters: the kernel parameters obtained from the two above iterations provide a search range for determining the final parameters. For example, $\sigma^{(1)}$ and $\sigma^{(2)}$ specify a range over which we will search for the final value of σ that can minimize the training error on set $\mathcal{T}$. Other kernel parameters can be jointly determined through the search.

In addition to determining kernel parameters, we also filter out the samples that have very similar values but different class labels. These samples are usually located in such region of the feature space that is difficult to classify and they can make the classification boundary very complex. Removing them from the training set helps improve the generalization capability of the classifier.

6.2.2 Overall Algorithm

The overall algorithm of our proposed receiver-side classification-based block concealment is summarized in Fig. 6.6. Below we explain a few additional details of the training and concealment processes.

Selection of training data We choose a set of training images that represent a variety of characteristics. Because of the spatial correlation in most natural images, we use about one fourth of blocks in the checkerboard pattern from each training image as candidates to form a training set. As discussed earlier, we further filter out the blocks where different concealment schemes do not give substantially different performance.

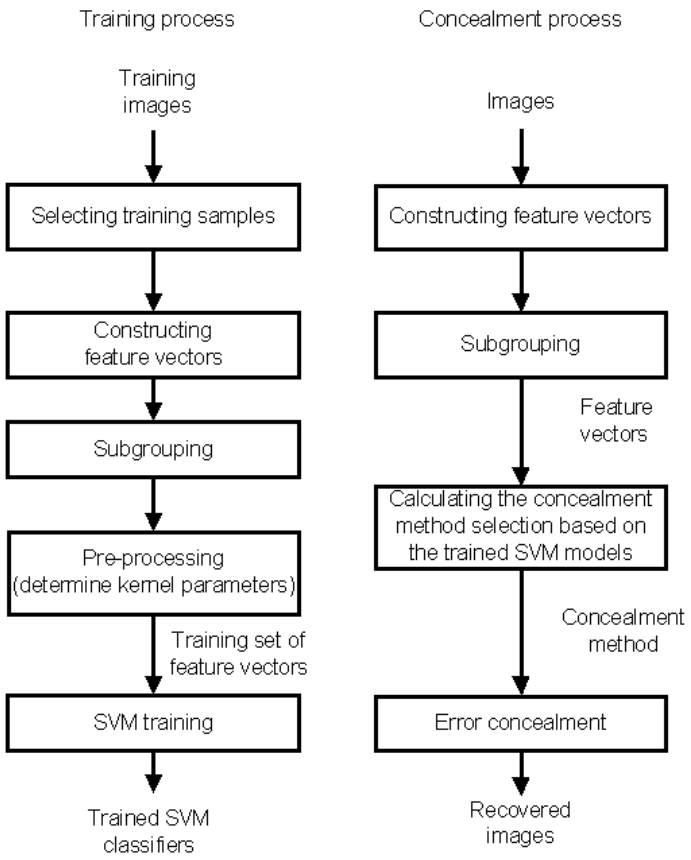


Figure 6.6: Block diagram of the proposed receiver-side classification-based concealment approach.

Construction of feature vectors Since different spatial block concealment techniques may use different sets of surrounding pixels, the feature vectors derived for classification should come from the union of the sets of pixels used by these techniques. For example, GSB often uses two surrounding layers to extract the geometric structure information, while OASI uses four surrounding layers to compute the interpolation coefficients. The classification region should therefore includes four surrounding layers of pixels. For block size of 8×8, 192 pixels are involved in classification.

While the pixels can be used directly as features, they often require a sophisticated kernel function to ensure separability and incur high computation complexity. We generate a more compact feature vector from pixel values using a similar approach as described in Section 6.1.3 and summarized as follows. We first partition the four surrounding layers of pixels into segments, as illustrated in Fig. 6.3(a). For the i^{th} segment of 4 pixels, the feature value v_i characterizes the smoothness of the segment and is computed as

$$v_i = \text{floor}[(\max\{p_k\} - \min\{p_k\} - s)/Q_v] + 1, \tag{6.9}$$

where $\{p_k\}$ are the pixels in the i^{th} segment, the floor function returns the largest integer less than or equal to the input. The two parameters s and Q_v control the sensitivity of the feature. We choose $s = 15$ and $Q_v = 50$ based on our experimental results. We then use these feature values to construct a feature vector. The ordering of features in the feature vector does not affect the performance of a trained SVM classifier since the kernel functions widely used in SVM classification are invariant with respect to the ordering of features. We have tried another scanning order in our previous experiments, which produced similar classification result [12].

Subgrouping As discussed earlier, to handle the nonlinearity of the class boundary, we divide the feature space into n subsets and train an SVM classifier for each subset. We use a simple empirical rule based on the number of nonzero values in a feature vector to perform the partitioning.

Pre-processing of training samples The feature vectors we used for training are divided into set $\mathcal{A}$ and $\mathcal{B}$. Each set includes images from all four representative categories mentioned before, namely, portraits, artificial images, natural scenery images, and rich texture images. We determine in this step the kernel parameters and training set using the approaches described in Section 6.2.1.

Concealment process After the training process is performed off-line, the parameters of trained SVM classifiers are stored in the receiver system. To conceal a corrupted image block, the receiver system use the same approach as in the training process to construct feature vector and identify to which subgroup the feature vector belongs to. The classification result will then determine which concealment scheme to use.

6.2.3 Experimental Results and Performance Analysis

In this section, we present the experimental results on the proposed block concealment method using receiver-side classification. We use the SVM^{light} toolkit [28] to accomplish this classification task. SVM^{light} is an implementation of SVM based on the optimization algorithm in [27].

A total of 15 images are used for training and 13 for testing, which are shown in Fig. 6.11 and Fig. 6.12. There are a total of 5,562 blocks in the training images

and 3,804 blocks in the test images having substantially different concealment performance by GSB and OASI. These blocks are used to evaluate the classification accuracy.

We first train a linear SVM using the 48-dimension feature vectors of all training blocks. The classification accuracy of this trained linear SVM on the test blocks is only 50.55%. The failure of this classification experiment indicates the high nonlinearity in the boundary of the two classes. We then examine the effects of various approaches in handling the nonlinearity. The simulation results of this exploration are shown in the first row of Table 6.4. We compare the cases of no subgrouping, 16-group subgrouping, and 48-group subgrouping. For these three cases, the kernel parameters are chosen that can provide the highest classification accuracy on three of the training images, "Lena," "Barbara," and "Bassharbor." We also consider the case of applying pre-processing with 48-group subgrouping for thorough selection of kernel parameters and filter out noisy samples, using the approaches described in Section 6.2.1. As shown in the table, subgrouping significantly improves the classification accuracy by more than 15%; and pre-processing and finer subgrouping can further improve the classification accuracy.

Based on results from the above exploration, we finally adopt 48 subgroups with pre-processing procedure for our training process, and examine the concealment performance of the proposed receiver-side classification-based scheme on the 13 8-bit gray-scaled test images. The classification accuracy for each subgroup ranges from 58.82% to 83.09%, and the overall classification accuracy is 67.11%. From the comparison of concealment results with that of GSB [76] and OASI [35] in Table 6.5, we can see that the classification-based method with a linear kernel has up to 0.84 dB gain when compared to the GSB method and up to 1.06 dB gain when compared

Table 6.4: Overall classification accuracy on the 13 test images

	1 group	Subgroup-16	Subgroup-48	Subgroup-48 with Pre-processing
Linear SVM	50.55%	65.96%	66.26%	67.11%
Radial SVM	65.54%	66.75%	67.17%	70.16%

to the OASI method.

We then train a radial basis kernel SVM to evaluate how well it handles the nonlinearity of training data. The pre-processing and subgrouping are also evaluated for this nonlinear kernel. As with the linear kernel, the radial basis kernel can also benefit from the pre-processing and finer subgrouping for improving the classification accuracy, although the improvement due to grouping is less significant on the radial basis kernel than on the linear kernel. This latter aspect is expected as the radial basis kernel has a good capability of handling the nonlinear classification boundary even without subgrouping. The classification accuracy for each group ranges from 60.00% to 80.53%, and the overall classification accuracy is 70.16%. As shown in Table 6.5, the classification-based method using the radial basis kernel SVM has up to 0.94 dB gain compared to the GSB method and up to 1.26 dB gain compared to the OASI method. The proposed scheme consistently outperforms the two prior algorithms on all test images. As an example, we show a portion of the "Nickel" image in Fig. 6.7, and we can see that the proposed concealment scheme provides better visual quality and leaves fewer artifacts.

It is worth noting that a radial basis kernel gives about 3% higher classification

accuracy than a linear kernel, under the same 48-group subgrouping and preprocessing procedure. The small improvement in classification accuracy, however, does not always translate into the improvement of concealment quality. For example, we can see from Table 6.5 that radial basis kernel provides slightly better concealment for some test images, while linear kernel is better for others. This is because the set of accurately classified blocks may be different by the two kernel techniques, and the quality gain on the slightly bigger set of accurately classified blocks may not always offset the quality loss on the falsely classified ones. On the other hand, we see that the classification-based schemes give consistently higher concealment quality than the two current state-of-the-art algorithms. With more accurate classification, the concealment quality can be further improved. Along the line of seeking for more accurate classification information, we are inspired by the growing importance of involving both sender and receiver in efficient and reliable visual communications. In the next section, we investigate what role the sender system can play in facilitating the classification-based concealment.

6.3 Block Concealment with Sender-Supplied Classification Information

The receiver-side classification algorithm proposed in Section 6.2 outperforms the conventional error concealment approaches. Coming with such benefit is the increase in computation complexity at receiver-side for performing classification. The increased complexity may pose a challenge for systems that have very limited computation resources and/or stringent real-time rendering constraints. If some parts of the concealment task could be moved to the sender side, it would help

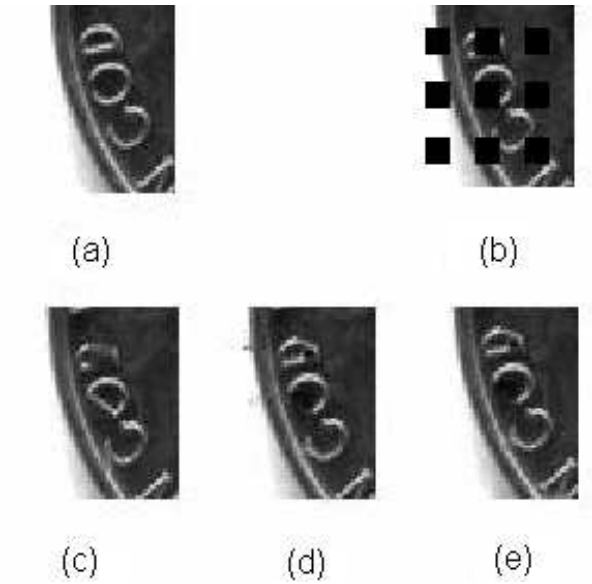


Figure 6.7: Visual quality comparison of three concealment schemes: (a) original image; (b) corrupted image; (c) recovered image using GSB; (d) recovered image using OASI; and (e) recovered image using the classification-based method.

Table 6.5: Comparison of concealment quality in PSNR (dB) of existing concealment schemes and the proposed receiver-side classification-based approaches. Type A: Natural; Type B: Portrait; Type C: Artificial; Type D: Texture.

Type	Name	GSB	OASI	Better-2	Linear Kernel	Radial Kernel
A	Fishingboat	30.93	31.10	32.28	31.36	31.64
	Goldhill	32.35	32.41	33.52	32.63	32.84
	Peppers	35.18	35.55	36.72	36.02	35.79
	Skylinearch	32.01	31.34	33.22	32.40	32.60
	Lochness	32.74	32.33	33.40	32.78	32.78
	Bellflower	33.27	33.70	35.57	34.12	34.21
	Brandyrose	39.47	39.27	40.42	39.86	39.80
	Lake	28.54	28.73	30.14	29.10	29.04
	F14	38.64	38.86	39.88	38.75	39.05
B	Elaine	35.17	35.93	36.35	35.85	35.96
	Couple	30.74	31.06	32.22	31.49	31.43
C	Nickel	29.05	28.55	30.53	29.33	29.58
D	Baboon	26.11	26.48	27.12	26.62	26.62

reduce the computation burden on the receiver side, as demonstrated in recent works [73, 74].

An important benefit of moving the classification task from a receiver to a sender is that it allows for an easy access of the perfect classification information. This is because the sender has full reference to the original, uncorrupted image, and can compare the concealment quality by various techniques to obtain the ground truth about which technique works better. The higher accuracy of the classification information can further improve the overall concealment quality upon what we have achieved in Section 6.2, which is an even more attractive advantage than the reduced receiver-side computation complexity.

In this section, we extend the classification-based concealment framework from a sender-driven perspective to design and evaluate error concealment schemes with sender-supplied classification information. We shall examine two main approaches to conveying the classification information from a sender to a receiver: one is by attaching the side information in the header, and the other is to embed the side information in the image signal using data hiding technique.

6.3.1 Conveying Classification Information by Attachment

A quite straightforward way to convey the classification information from the sender to the receiver is to transmit the information along with the image, for example, in the image header. The side information requires extra bandwidth. Therefore, this attaching approach may be appropriate depending on the application and the image size. An alternative approach to avoid the increase in bandwidth is to encode the image at a lower rate to spare room for the side information. This would reduce the image quality, leading to a similar tradeoff as in the data embedding

approach to be discussed in the next subsection.

We present the system block diagram of the sender-side attaching scheme in Fig. 6.8. On the sender side, in addition to encoding an image as usual, the system would perform the following tasks:

1. Perform error concealment on each block or some selected blocks using multiple error concealment methods.

2. Compare the quality of the images obtained by these concealment methods, and classify each block according to the winning technique.

3. Encode the classification information for each block, possibly using lossless compression techniques.

4. Attach the classification information to the compressed image bit stream.

On the receiver side upon detecting the corrupted blocks, the receiver will extract the classification information from the received stream and use this side information to select the appropriate method for concealing each corrupted block. We can further apply forward error correction coding with appropriate strengths to protect the image stream and the side information.

Regarding the detailed encoding method for side information, we denote the side information for the GSB concealment method as "0," and that for OASI as "1." The side information for all blocks can be put together as a binary sequence. Recall that GSB concealment has lower computation complexity than OASI. So as before, we choose the error concealment technique with lower computation complexity for the blocks where the performance of the two concealment methods are not significantly different. This also helps give a long run of "0" in the side-information

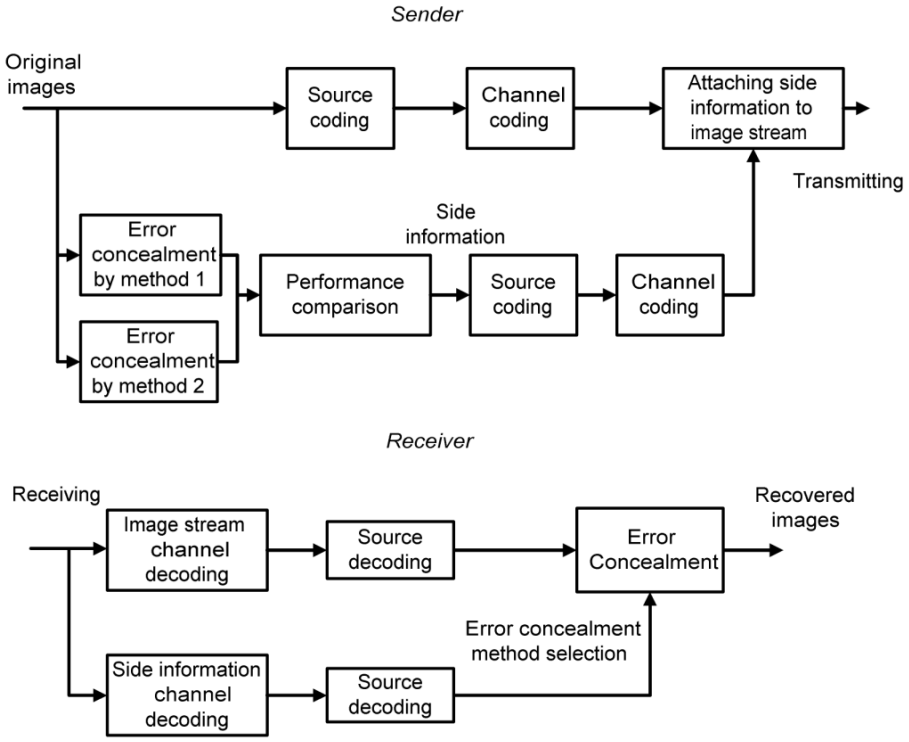


Figure 6.8: Block diagram of the sender-side attaching approach.

encoding. We then apply run-length coding and arithmetic coding to compress the binary sequence of classification information.

We have seen that the attaching scheme trades additional bandwidth for improved concealment quality. The trade-off can be adjusted as follows. For each block, the performance of each algorithm ($P1$ and $P2$) is calculated according to (6.1). The binary-valued side information L for the block is determined by

$$L = \begin{cases} 1, & \text{if} \quad P1 - P2 > \Delta_{th} \\ 0, & \text{otherwise} \end{cases}, \tag{6.10}$$

where Δ_{th} is a threshold. An experiment with different settings of Δ_{th} is performed on the JPEG compressed "Lena" image with quality factor $Q = 80\%$, where the image size is 512×512 and the JPEG file size is 303,072 bits. As shown in Fig. 6.9, the larger Δ_{th} we choose, the lower PSNR we get. On the other hand, since more blocks are labeled as "0" with a larger Δ_{th}, compressing the classification information using run-length coding and arithmetic coding will achieve a higher compression ratio. The results in Fig. 6.9 shows that when Δ_{th} is around 96, the gain in error concealment quality is significant, yet the additional bandwidth for classification side information is quite moderate and only about one percent of the image file size. Thus we use this value to evaluate the overall concealment quality.

The simulation results of the attaching scheme are listed in Table 6.6. The results suggest that our proposed concealment scheme by attaching classification information outperforms each individual receiver-side concealment approach. The error concealment quality can be improved by about $1 \sim 2$ dB when compared to the better one between the two individual methods. Readers may notice that the attaching scheme has 0 dB gain on the "Circletrain" image when compared to GSB. As shown in Fig. 6.11, this artificial image has uniform background and smooth

Table 6.6: Performance evaluation of the sender-side attaching approach. Type A: Natural; Type B: Portrait; Type C: Artificial; Type D: Texture.

Type	Name	JPEG file size (bytes)	Side info. size (bytes)	Gain over GSB (dB)	Gain over OASI (dB)
A	Bassharbor	50,867	368	0.52	1.14
	Blueflower	53,528	495	0.87	0.90
	House	46,975	361	1.28	1.26
	Newyork	73,830	436	0.89	0.67
	Operahouse	48,666	365	1.09	0.99
	Papermachine	41,773	285	1.95	2.07
	Watch	41,773	293	1.23	1.09
B	Lena	37,884	287	0.99	0.93
	Barbara	50,867	424	2.21	1.12
	Kid	30,791	257	1.12	1.08
	Man	61,810	431	0.80	0.79
C	Circletrain	15,709	124	0	11.19
	Tulip	48,641	437	1.45	1.68
	Waterfall	44,734	292	0.93	0.75
D	Bear	26,089	280	1.32	1.12

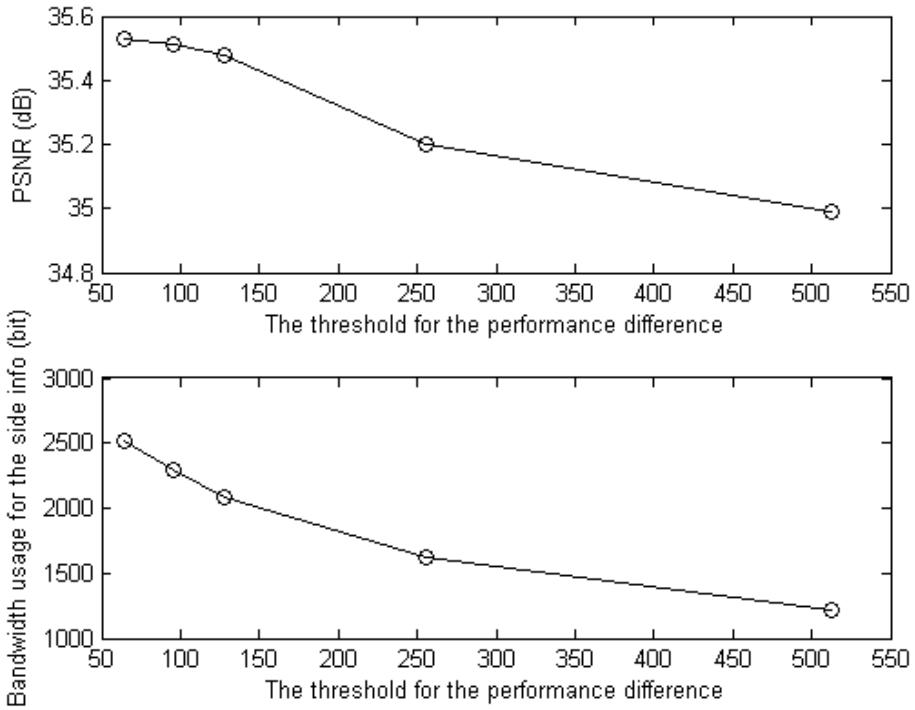


Figure 6.9: Relation of the threshold Δ_{th} versus the concealment quality and the bandwidth required for side information, respectively, when applying the sender-side attaching approach on the "Lena" image.

edges. GSB gives better concealment quality in terms of PSNR for every recovered blocks, so we cannot get any improvement compared to GSB.

6.3.2 Conveying Classification Information by Embedding

Although the attaching scheme has excellent performance, the additional bandwidth for side information may not be available or too pricey in some systems. Recoding the image part to a slightly lower rate requires a non-trivial amount of computation complexity to ensure that the total bandwidth of the image plus the side information is unchanged. A viable alternative to convey side information with little additional bandwidth is embedding it in the image. More specifically, we embed 1-bit classification information of a block into its neighboring block. The embedding will be incorporated in the visual communication system along with interleaved packetization mentioned at the beginning of the chapter, so that the neighboring blocks are packed into different packets. In such a way, it is unlikely for a block and its neighbor holding its classification information to be corrupted simultaneously. As we shall see later in this subsection, the embedding in neighbor block has additional advantage when dealing with smooth blocks. We summarize the system block diagram in Fig. 6.10 and explain a few details of embedding below.

As can be seen from the previous subsection, the amount of classification information is on the order of a couple of thousands bits, which calls upon an embedding technique with quite high embedding rate. Unlike many copyright protection applications, there is no major adversary to circumvent the embedded data in the error concealment application, where the side information helps improve the performance of image communications [70, 10, 58]. The quantization based data embedding is a common choice to meet these requirements [71].

We use a simple version of quantization embedding, known as the *odd-even embedding* technique [72], to embed the classification information into the image. To avoid a substantial impact on the compression size and the visual quality of the image, the classification information for each block is embedded into the last quantized non-zero DCT coefficient in the zig-zag scan order. The coefficient is forced to be even if we want to embed "0," or odd if to embed "1," and the embedding tries to make minimum necessary changes to enforce such relation. If all the quantized AC coefficients in a block are zero, which we would encounter for smooth blocks, we will not make any changes on the coefficients. In this case, the receiver would consider a "0" is embedded in the block based on the above-mentioned rules, and apply the concealment technique of lower computation complexity (i.e., GSB) for the corrupted block. Such an arrangement works well in practice. This is because GSB usually performs better for blocks with relatively "flat" surrounding; in the mean time, the characteristics of nearby blocks are likely to be similar and can be fully exploited by neighborhood embedding presented earlier, where classification information is embedded into neighboring block.

The experimental results of the embedding scheme are shown in Table 6.7. The improvement of concealment quality on most images is significant: we have a $0.14 \sim 1.94$ dB gain compared to GSB, and $0.24 \sim 1.14$ dB gain compared to OASI. For most images, GSB performs better on some blocks, and OASI performs better on some other blocks. As such, the quality degradation introduced by the embedding procedure is overcome by the substantial concealment gain compared to either GSB or OASI alone. An interesting exception appears on the "Circletrain" image. Different from other images, GSB is the better selection for all blocks in the "Circletrain" image and the concealed quality is very high (The PSNR value in dB

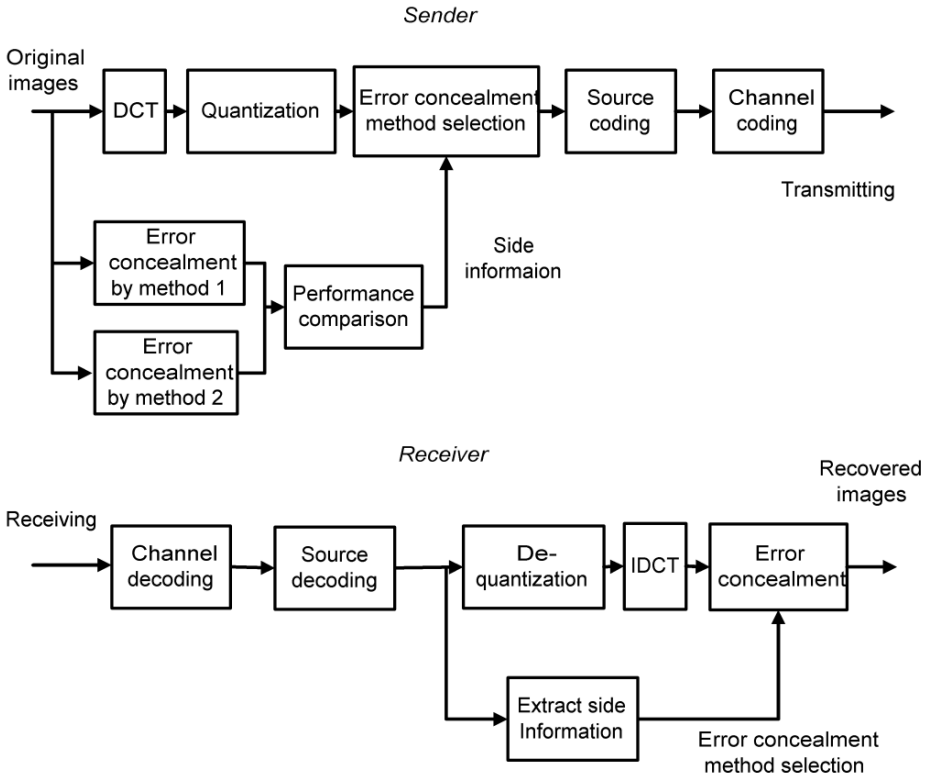


Figure 6.10: Block diagram of the sender-side embedding approach.

is in the high forties). The sender-supplied classification information thus provides no gain when compared to using GSB alone. On the other hand, the embedding technique inevitably introduces a moderate amount of quality degradation. As the result, for the "Circletrain" image, the embedding scheme achieves a net loss of 2.8 dB in PSNR compared to GSB, although little visual difference could be visible at such high PSNR levels. In comparison with OASI, the gain over OASI is over 8 dB and is much more noticeable.

6.4 Comparisons and Discussions

In the previous two sections, we have proposed three classification-based error concealment schemes to improve the concealment quality. Among the three schemes, one performs classification on the receiver side using features derived from the survived pixels surrounding a corrupted block and an SVM classifier, and the other two schemes convey the sender-supplied classification information to receiver by attaching and embedding, respectively. As we can see from Tables 6.5, 6.6, and 6.7, they all improve the concealment quality quite substantially. In this section, we compare the three schemes, discuss their advantages and shortcomings, and identify the application scenarios that each scheme is suitable for. We also discuss a few directions for further extension and generalization.

We first compare the quality of concealed images by these three schemes and show the results in Table 6.8. For each image, we use the uncorrupted JPEG compressed version with a quality factor of 80% as reference. Since the attaching scheme provides the ground-truth of concealment technique selection to the receiver, it gives the highest concealment quality among the three schemes. The improvement over the individual concealment schemes is in the range of $0.5 \sim 1.5$ dB. While the

Table 6.7: Performance evaluation of the sender-side embedding approach. Images are in the JPEG format with quality factor $Q = 80\%$. Type A: Natural; Type B: Portrait; Type C: Artificial; Type D: Texture.

Type	Name	PSNR of image after embedding	Gain over GSB (dB)	Gain over OASI (dB)
A	Bassharbor	41.89	0.14	0.76
	Blueflower	41.73	0.77	0.80
	House	42.01	1.00	0.98
	Newyork	38.25	0.74	0.52
	Operahouse	40.57	0.63	0.53
	Papermachine	42.42	1.02	1.14
	Watch	42.82	0.74	0.60
B	Lena	43.21	0.30	0.24
	Barbara	42.25	1.94	0.85
	Kid	43.16	0.60	0.56
	Man	39.48	0.49	0.48
C	Circletrain	47.36	-2.80	8.39
	Tulip	42.31	0.78	1.01
	Waterfall	40.47	0.62	0.44
D	Bear	43.91	0.63	0.43

Table 6.8: Comparison of concealment quality in PSNR (dB) by the receiver-side and sender-side approaches. Images are in the JPEG format with quality factor $Q = 80\%$.

Image Type	Image Name	GSB	OASI	Receiver-side Classification	Sender-side Embedding	Sender-side Attaching
Natural	Fishingboat	30.81	30.87	31.03	31.55	32.02
Portrait	Elaine	35.47	35.18	35.43	35.84	36.22
Artificial	Nickel	28.48	28.41	28.71	29.40	29.93
Texture	Baboon	26.02	26.19	26.25	26.45	26.74

embedding scheme also provides the ground-truth of most blocks to receiver (except for some very smooth blocks), its performance is lower than the attaching scheme by about $0.3 \sim 0.5$ dB. The small quality loss is due to the distortion introduced by embedding, a price paid for sending side information without additional bandwidth. The receiver-side classification scheme has the smallest improvement over individual scheme because the classification result at the receiver is not always accurate.

In addition to the quality of concealed image, other important issues include computation complexity, bandwidth usage, and complexity associated with overall system deployment. The receiver-side classification-based error concealment requires neither side information to be sent nor any special involvement of a sender. It can be therefore integrated in a standard-compliant coding system. The training involves a large amount of computation but can be performed off-line. A moderate amount

of run-time computation power is required from the receiver to extract features and feed them into a trained SVM classifier to determine which concealment scheme to use, and this is done only for corrupted blocks. As the classification results are not always perfect and depend heavily on the generalization capability of the classifier, the concealment performance may vary substantially from one image to another. This scheme is suitable for applications where there is limited design flexibility on the sender side.

The schemes with sender-supplied classification information provides more *proactive* protection. They require a significant amount of computation power and cooperation on the sender side to perform concealment, provide ground-truth on the concealment scheme to use for *every* block, and encode or embed the classification information with the image. The attaching scheme requires additional bandwidth to deliver the ground-truth of classification. After such an attachment, the resulting media stream may not be standard-compliant. In contrast, the embedding scheme can maintain standard compliance of the resulting media stream. This is at an expense of minor reduction of the perceptual quality in the transmitted image, even when the transmission is free from error. On the other hand, the more accurate, sender-supplied classification information provides substantial improvement in concealment quality and also eliminate the computation need on the receiver side for classification. These schemes are suitable for applications with powerful sender and simple receiver, and for scenarios where the visual data is encoded once but delivered and consumed by many users.

The spatial concealment schemes investigated in this chapter can be used for both image and video transmissions. They can be applied to each corrupted video frame, and can be used in conjunction with other temporal concealment methods [18,

33]. The schemes that maintain standard-compliance of the transmitted video, such as the receiver-end classification and the embedding schemes, allows image/video to be handled by a number of existing visual communication systems that support the standard, with little additional changes to the system.

In addition to conveying side information to facilitate concealment, data embedding can also be used for detecting corrupted blocks [16]. For this error detection purpose on each block, the parity information or some known patterns should be embedded *inside* the corresponding block. The receiver will check the correctness of the parity or the integrity of the patterns to determine whether the block is corrupted. On the other hand, the side information of a block for facilitating its concealment must be stored *outside* that block, as seen in the algorithm presented in the previous section.

We have so far assumed that the block damage is isolated (i.e., all neighboring blocks of the damaged one are correctly received). Since consecutive block damage is a challenge to most error concealment techniques, interleaving techniques have been suggested in packetization to avoid packing neighboring blocks together [67] [74]. As such, consecutive block losses rarely happen at a moderate loss rate. In case when there remain some consecutive block losses, both GSB and OASI techniques have been demonstrated to handle a small number of consecutive blocks [76, 35]. The classification can also be extended to cope with this case, for example, to incorporate the loss of two horizontal or vertical neighboring blocks by training more classifiers. Since what we have proposed is a general framework, it can be further extended in several directions. For example, we can incorporate other concealment techniques and the total number of candidate techniques can be more than two.

Bassharbor (512x512)

Blueflower (512x512)

House (512x512)

Newyork (512x512)

Operahouse (512x512)

Papermachine (512x512)

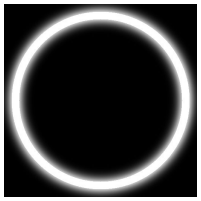
Watch (512x512)

Lena (512x512)

Barbara (512x512)

Kid (480x480)

Man (512x512)



Circletrain (512x512)

Tulip (512x512)

Waterfall (512x512)

Bear (384x384)

Figure 6.11: The 15 8-bit gray-scaled images are used for training in classification. The image sizes are listed in parentheses after the image names.

Fishingboat (512x512)

Goldhill (512x512)

Peppers (512x512)

Skylinearch (400x400)

Lochness (512x512)

Bellflower (512x512)

Brandyrose (512x512)

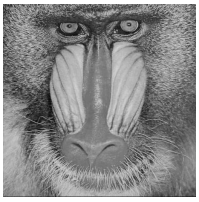
Lake (512x512)

F14 (496x496)

Elaine (512x512)

Couple (512x512)

Nickel (256x256)



Baboon (512x512)

Figure 6.12: The 13 8-bit gray-scaled images are used for testing in the classification-based concealment. The image sizes are listed in parentheses after the image names.

6.5 Chapter Summary

In this chapter, we present a new, classification-based spatial error concealment framework for images. Our proposed framework takes advantages of state-of-the-art concealment techniques and adaptively selects the best suitable one for each corrupted block. Using the new framework, we have designed concealment schemes outperforming the current state-of-the-art algorithms in terms of the error concealment quality on a diverse set of images. The proposed framework also allows the computation power to be spent more strategically by using a computation demanding algorithm only when it can significantly improve the recovered image quality.

Chapter 7

Conclusions and Future Perspectives

This dissertation provides the frameworks to explore heterogeneities in error-resilient visual communications, such as the characteristics heterogeneity amongst the content, the different channel condition of communication links, and the different demand requirement with which multimedia streams are consumed. The challenges in heterogeneous visual communications urge the need of integrating error-resilient techniques with other techniques, such as resource allocation or classification. The proposed FEC and resource allocation integrated approach and classification-based error concealment approach can successfully adapt compressed streams of visual content with different characteristics to be resilient to heterogeneous network conditions. More specifically, the main contributions of this dissertation are:

First, we have developed a framework of distributed multi-point video conferencing system over packet erasure channels. As presented in Chapter 2, the framework handles the error protection and resource allocation of multiple video streams in a distributed manner. A PDMA-based error protection scheme performed in each video stream combiner is proposed to explore the multi-stream heterogeneity.

In Chapter 3, PDMA-based error protection scheme is modeled as an opti-

mization problem to minimize the maximal expected video distortion among all aggregated streams. A fast algorithm is proposed to provide the optimal solution. The simulation results show that our proposed multi-stream video aggregation and error protection scheme has significant gain over traditional multi-stream error protection schemes.

In order to deliver video streams to end users with different preferred quality, we investigate an approach that adapts multi-stream aggregation to user preference heterogeneity in Chapter 4. We propose a consensus algorithm to perform resource allocation based on user preference.

In Chapters 2–4, each hop solves its own aspect of FEC problem based on resource availability and channel reliability of a single hop. There is no overall consideration of multiple sequential hops' resource and channel condition. We furthermore find out that, given a limited bandwidth, the method of applying optimal FEC and bandwidth allocation locally in one node may not achieve the optimal result for multi-hop multi-stream video aggregation. The fundamental reason is that the video content loss in the previous hop cannot be recovered in the following hop no matter how much bandwidth is available and how strong FEC is applied. A method is then investigated in Chapter 5 to accommodate multi-hop heterogeneity in our FEC and resource allocation integrated approach. This method explores multi-hop awareness and multi-stream heterogeneity. It performs optimization in FEC and resource allocation to minimize the maximum distortion across all video streams delivered to all end users. Our simulation results show that the proposed multi-hop awareness method has significant gain in terms of the perceptual quality of delivered video streams comparing to the method without multi-hop awareness,.

We address heterogeneity issue involved in error concealment in Chapter 6.

Due to the large variation of image characteristics, different concealment approaches are necessary to accommodate different nature of the lost image content. We propose using classification to integrate state-of-the-art error concealment techniques. The proposed approach takes advantages of multiple concealment algorithms and adaptively selects the suitable algorithm for each damaged image area. With growing awareness that the design of sender and receiver systems should be jointly considered for the efficient and reliable visual communications, we proposed a set of classification-based block concealment schemes, including receiver-side classifying, sender-side attaching, and sender-side embedding. Our experimental results provide extensive performance comparisons and demonstrate that the proposed classification-based error concealment approaches outperform the conventional approaches.

In summary, this dissertation presents promising frameworks for heterogeneity exploration of error-resilient visual communications. The research work in this dissertation can lead to designing a platform for large-scale real-time video streaming over heterogeneous networks.

In this dissertation, we mainly use effective resource allocation and classification principles to achieve error resiliency in visual communications involving many kinds of heterogeneities. For the future directions of our research work, we are interested in studying if applying operations of introducing redundancy across multiple multimedia streams can effectively accommodate multi-stream heterogeneity in FEC. For example, digital fountain code [5] shows its promising advantages in asynchronous, one-to-many, and on-demand applications, such as file download or movie download. Intuitively, if we interleave the source symbols from different streams with a certain pattern first, allow the operation of introducing redundancy

(e.g. XOR) in digital fountain code to be applied between symbols from different streams, there may be advantages for error resiliency in concurrent downloading multiple video or data streams.

We also have further interests in investigating a combined approach of FEC and error concealment for error-resilient visual communications. In Section 6.3 of this dissertation [13], we discussed the method of embedding certain information into multimedia streams to facilitate error recovery at the decoder. This method can be viewed as adding redundant information to transmitted multimedia streams as well. For future perspectives of our research work, some questions arise: What is the performance of this embedding-based method compared to traditional FEC? Are there certain characteristics of visual content that favor either of these two error-resilient techniques? Can we intelligently apply these two techniques at the same time to achieve optimal error-resilient result? These questions could lead to interesting research in error-resilient visual communications.

BIBLIOGRAPHY

[1] E. Amir, S. McCanne and R. Katz. An Active Service Framework and Its Application to Real-Time Multimedia Transcoding. *ACM SIGCOMM*, pages 178-189, Sep. 1998.

[2] S. Appadwedula, D. L. Jones, K. Ramchandran and L.R. Qia. Joint source-channel matching for wireless image transmission. *Proc. IEEE Int'l Conf. Image Processing*, vol. 2, pages 137-141, Oct. 1998.

[3] M. Bertalmio, L. Vese, G. Sapiro and S. Osher. Simultaneous Structure and Texture Image Inpainting. *IEEE Trans. Image Processing*, 12(8):882-889, Aug. 2003.

[4] C. J. C. Burges. A Tutorial On Support Vector Machines for Pattern Recognition. *Data Ming and Knowledge Discovery*, 2(2):121-167, June 1998.

[5] J. W. Byers, M. Luby, and M. Mitzenmacher. A Digital Fountain Approach to Asynchronous Reliable Multicast. *IEEE Journal on Selected Areas in Communications*, 20(8):1528-1540, Oct., 2002.

[6] J. Cai and C. W. Chen. FEC-Based Video Streaming over Packet Loss Networks with Pre-Interleaving. *IEEE Int'l Conf. on Information Technology: Coding and Computing*, pages 10-13, April 2001.

[7] H. Cai, B. Zeng, G. Shen and S. Li. Error-Resilient Unequal Protection of Fine Granularity Scalable Video Bitstreams. *IEEE Int'l Conf. on Communications*, vol. 3, pages 1303-1307, June 2004.

[8] R.N. Chang, Z.-Y. Shae, X. Gu and K. Nahrstedt. An Overlay Based QoS-Aware Voice-over-IP Conferencing System. *IEEE Int'l Conf. on Multimedia and Expo*, vol. 3, pages 2111-2114, Jun. 2004.

[9] T. P.-C. Chen and T. Chen. Error Concealment Aware Rate Shaping for Wireless Video Transport. *EURASIP Signal Processing: Image Communication*, pages 889-905, 2003.

[10] M. Chen and G. Bailey. Image Quality Assessment Using Data Hiding for Performance Evaluation of Visual Communication Networks. *Int'l working Conf.*

on Performance Modelling and Evaluation of Heterogeneous Networks, July 2005.

[11] M. Chen, G.-M. Su and M. Wu. Robust Distributed Multi-Point Video Conferencing over Error-Prone Channels. *IEEE Int'l Conf. on Multimedia and Expo*, pages 1149-1152, July 2006.

[12] M. Chen, M. Wu and Y. Zheng. Classification-Based Spatial Error Concealment For Images. *Proc. IEEE Int'l Conf. Image Processing*, pages 675-678, Sep. 2003.

[13] M. Chen, Y. Zheng, and M. Wu, Classification-Based Spatial Error Concealment for Visual Communications. *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 13438, 17 pages, 2006.

[14] M. Chen, G.-M. Su and M. Wu. Distributed Multi-Point Video Conferencing over Packet Erasure Channels. Submitted to *IEEE Trans. on Multimedia*, April 2007.

[15] M. Chen, G.-M. Su and M. Wu. Robust Multi-strem Video Aggregation with Multi-hop Awareness. In preparation.

[16] M.-H. Chen, Y. He and R. L. Lagendijk. A Fragile Watermark Error Detection Scheme For Wireless Video Communications. *IEEE Trans. Multimedia*, 7(2):201-211, April 2005.

[17] T.-C. Chen, S.-M. Lei and M.-T. Sun. Video Bridging Based on H.261 Standard. *IEEE Trans. Circuits Syst. Video Technol.*, 4(4):425-437, Aug. 1994.

[18] C. Dovrolis, D. Tull and P. Ramanathan. Hybrid Spatial/Temporal Loss Concealment for Packet Video. *Proc. Int'l Packet Video Workshop*, April 1999.

[19] R. O. Duda, P. E. Hart and D. G. Stork. Pattern Classification. *John Wiley and Sons Inc.*, 2001.

[20] K.-T. Fung, Y.-L. Chan and W.-C. Siu. Low-Complexity and High-Quality Frame-Skipping Transcoder for Continuous Presence Multipoint Video Conferencing. *IEEE Trans. on Multimedia*, 6(1):31-46, Feb. 2004.

[21] E. Gabrielyan. Fault-Tolerant Real-Time Streaming with FEC thanks to Capillary Multi-Path Routing. *Proc. IEEE Int'l Conf. on Communications, Circuits and Systems*, vol. 3, pages 1497-1501, June 2006.

[22] V. K. Goyal and J. Kovacevic. Generalized Multiple Description Coding with Correlating Transforms. *IEEE Trans. on Multimedia*, 47(6):2199-2224, Sep. 2001.

[23] X. Gu and K. Nahrstedt. Distributed Multimedia Service Composition with Statistical QoS Assurances. *IEEE Trans. on Multimedia*, 8(1):141-151, Feb. 2006

[24] S. Gringeri, R. Egorov, K. Shuaib, A. Lewis and B. Basch. Robust Compression and Transmission of MPEG-4 Video. *ACM Multimedia*, pages 113-120, Aug. 2000.

[25] A. Harris, C. Sengul, R. Kravets and P. Ratanchandani. Energy-Efficient Transmission Policies for Multimedia in Multi-hop Wireless Networks. *IEEE Int'l Conf. on Mobile and Wireless Communication Networks*, Oct. 2004.

[26] T. Joachims. Text Categorization with Support Vector Machines: Learning with Many Relevant Features. *Proc. European Conf. Machine Learning*, pages 137-142, April 1998.

[27] T. Joachims. Making large-scale SVM learning practical, in *Advances in Kernel Methods Support Vector Learning*, pages 169184 B. Schlkopf, C. Burges, and A. Smola, Eds. MIT Press, Cambridge, MA, 1999.

[28] T. Joachims. SVM^{light} Support Vector Machine V5.00, *http://svmlight.joachims.org/*, 2002.

[29] K. Jung, J. Chang and C. Lee. Error Concealment Technique Using Projection Data for Block-based Image Coding. *SPIE Conf. Visual Comm. & Image Processing*, pages 1466-1476, Sep. 1994.

[30] A.K. Katsaggelos, Y. Eisenberg, F. Zhai, R. Berry and T.N. Pappas. Advances in Efficient Resource Allocation for Packet-Based Real-Time Video Transmission. *Proceedings of the IEEE*, 93(1):135-147, Jan. 2005.

[31] J. Kim, R. M. Mersereau and Y. Altunbasak. Distributed Video Streaming Using Multiple Description Coding and Unequal Error Protection. *IEEE Trans. on Multimedia*, 14(7):849-861, July 2005.

[32] J. Kim, R. M. Mersereau and Y. Altunbasak. Error-Resilient Image and Video Transmission over the Internet Using Unequal Error Protection. *IEEE Trans. on Multimedia*, 12(2):121-131, Feb. 2003.

[33] Y. C. Lee, Y. Altunbasak and R. M. Mersereau. A Temporal Error Concealment Method for MPEG Coded Video Using a Multi-Frame Boundary Matching Algorithm. *Proc. IEEE Int'l Conf. Image Processing*, pages 990-993, Oct. 2001.

[34] Y.-J Lee, T.-B. Lim, Y.-S. Kim and S.-P. Lee. Development of a Seamless Data Streaming System Based on User Preference and Device Information. *IEEE Int'l Conf. on Software Engineering Research, Management and Applications*, pages 260-267, Aug. 2006.

[35] X. Li and M. T. Orchard. Sequential Error-Concealment Techniques Using Orientation Adaptive Interpolation. *IEEE Trans. on Circuits and Systems for Video Technology*, 12(10):857-864, Oct. 2002.

[36] D. Li, Q. Zhang, C-N. Chuah and S. J. B. Yoo. Error Resilient Concurrent Video Streaming over Wireless Mesh Networks. *Int'l Packet Video Workshop*, April 2006.

[37] Y. J. Liang, J. G. Apostolopoulos and B. Girod. Model-Based Delay-Distortion Optimization for Video Streaming Using Packet Interleaving. *IEEE Asilomar Conference on Signals, Systems, and Computers*, vol. 2, pages 1315-1319, Nov. 2002.

[38] C.-W. Lin, Y.-C. Chen and M.-T. Sun. Dynamic Region of Interest Transcoding for Multipoint Video Conferencing. *IEEE Trans. Circuits Syst. Video Technol.*, 13(10):982-992, Oct. 2003.

[39] S. Lin and D. J. Costello. *Error Control Coding: Fundamentals and Applications*. Prentice Hall Inc., 1983.

[40] W.-H. Lin and A. Hauptmann. News Video Classification using SVM-based Multimodal Classifiers and Combination Strategies. *Proceedings of ACM International Conference on Multimedia*, pages. 323-326, Dec. 2002.

[41] J.-L. Lin, W.-L. Hwang and S.-C. Pei. SNR Scalability Based on Bitplane Coding of Matching Pursuit Atoms at Low Bit Rates: Fine-Grained and Two-Layer. *IEEE Trans. Circuits Syst. Video Technol.*, 15(1):3-14, Jan. 2005.

[42] Y. Liu and Y. Li. Error Concealment of Digital Images Using Data Hiding. *Proc. IEEE DSP Workshop*, Oct. 2000.

[43] R. Ma and J. Ilow. Regenerating Nodes for Real-Time Transmissions in Multi-Hop Wireless Neiworks. *Proc. IEEE Int'l Conf. on Local Computer Networks*, pages 378-384, Nov. 2005.

[44] A. Majumda, D. G. Sachs, I. V. Kozintsev, K. Ramchandran and M.M. Yeung. Multicast and Unicast Real-Time Video Streaming over Wireless LANs. *IEEE Trans. Circuits Syst. Video Technol.*, 12(6):524-534, Jun. 2002.

[45] C. Mayer, H. Crysandt and J.-R. Ohm. Encoding Multimedia Presentations for User Preferences and Limited Environments. *IEEE Int'l Conf. on Multimedia and Expo*, pages 165-168, July 2003.

[46] C. Mayer, H. Crysandt and J.-R. Ohm. Bit Plane Quantization for Scalable Video Coding . *SPIE Conf. Visual Comm. & Image Processing*, pages 1142-1152, Jan. 2002.

[47] S. McCanne, M. Vetterli and V. Jacobson. Low-Complexity Video Coding for Receiver-Driven Layered Multicast. *IEEE Journal on Selected Areas in Communications*, 15(6):983-1001, Aug. 1997.

[48] P. Mehra and A. Zakhor. TCP-Based Video Streaming Using Receiver-Driven Bandwidth Sharing. *Packet Video*, April 2003.

[49] A.E. Mohr, E.A. Riskin and R.E. Ladner. Generalized Multiple Description Coding Through Unequal Loss Protection. *IEEE Int'l Conf. on Image Processing*, vol. 1, pages 411-425, Oct. 1999.

[50] A. E. Mohr, E. A. Riskin and R. E. Ladner. Unequal Loss Protection: Graceful Degradation of Image Quality over Packet Erasure Channels Through Forward Error Correction. *IEEE Journal on Selected Areas in Communications*, 18(7):819-828, April 2000.

[51] W.T. Ooi and R. van Renesse. Destributing Media Transformation Over Multiple Media Gateways. *ACM Multimedia*, vol. 9, pages 159-168, Sep. 2001.

[52] R. Puri, K.-W. Lee, K. Ramchandran and V. Bharghavan. Forward Error Correction FEC Codes Based Multiple Description Coding for Internet

Video Streaming and Multicast. *Signal Processing: Image Communicatioin*, 16(8):745-762, May 2001.

[53] H.M. Radha, M. van der Schaar and Y. Chen. The MPEG-4 Fine-Grained Scalable Video Coding Method for Multimedia Streaming over IP. *IEEE Trans. on Multimedia*, 3(1):53-68, Mar. 2001.

[54] H. Radha and M. Wu. Overlay and Peer-to-Peer Multimedia Multicast with Network-Embedded FEC. *Proc. IEEE Int'l Conf. Image Processing*, vol. 3, pages 1747-1750, Oct. 2004.

[55] Y. Shan, I. V. Bajic, S. Kalyanaraman and J. W. Woods. Overlay Multi-Hop FEC Scheme for Video Streaming over Peer-to-Peer Networks. *Proc. IEEE Int'l Conf. Image Processing*, vol. 5, pages 3133-3136, Oct. 2004.

[56] Y. Shan, I. V. Bajic, S. Kalyanaramana and J. W. Woods. Overlay Multi-Hop FEC Scheme for Video Streaming. *Elsevier Journal on Signal Processing: Image Communications, Special Issue on Video Networking*, 20(8):710-727, May 2005.

[57] A. Smola, P. Bartlett, B. Schlkopf and D. Schuurmans. Advances In Large Margin Classifiers, MIT Press, Cambridge, MA, 1999.

[58] J. Song and K.J.R. Liu. A Data Embedded video coding scheme for error-prone channels. *IEEE Trans. Image Processing*, 3(4):415-423, Dec. 2001.

[59] V.M. Stanković, R. Hamzaoui and Z. Xiong. Real-Time Error Protection of Embedded Codes for Packet Erasure and Fading Channels. *IEEE Trans. on Circuits and Systems for Video Technology*, 14(8):1064-1072, Aug. 2004.

[60] H. Sun and W. Kwok. Concealment of Damaged Block Transform Coded Images Using Projections onto Convex Sets. *IEEE Trans. Image Processing*, 4(4):470-479, April 1995.

[61] G.-M. Su, M. Chen and M. Wu. Cross-Path PDMA-Based Error Protection for Streaming Multiuser Video over Multiple Paths. *Proc. IEEE Int'l Conf. Image Processing*, pages 21-24, Oct. 2006.

[62] G.-M. Su and M. Wu. Efficient Bandwidth Resource Allocation for Low-Delay Multiuser Video Streaming. *IEEE Trans. Circuits Syst. Video Technol.*, 15(9):1124-1137, Sep. 2005.

[63] M.-T. Sun, A.C. Loui and T.-C. Chen. A Coded-Domain Video Combiner for Multipoint Continuous Presence Video Conferencing. *IEEE Trans. Circuits Syst. Video Technol.*, 7(6):855-863, Dec. 1997.

[64] W.-T Tan and A. Zakhor. Video Multicast Using Layered FEC and Scalable Compress. *IEEE Trans. Circuits Syst. Video Technol.*, 11(3):524-534, Mar. 2001.

[65] V. Vapnik. The Nature of Statistical Learning Theory. Springer-Verlag Publisher, 1995.

[66] N. Wakamiya, T. Yamashita, M. Murata and H. Miyahara. Integrated Resource Allocation Scheme for Real-Time Video Multicast. *IEEE Global Telecommunications Conference*, vol. 2, pages 1455-1459, Nov. 2002.

[67] Y. Wang and Q.-F. Zhu. Error Control and Concealment for Video Communication: A Review. *Proc. IEEE*, 86(5):974-991, May 1998.

[68] Y. Wang, Q.-F. Zhu and L. Shaw. Maximally Smooth Image Recovery in Transform Coding. *IEEE Trans. Commun.*, 41(10):1544-1551, Oct. 1993.

[69] W. Wilson and H. Sun. Multi-Directional Interpolation for Spatial Error Concealment. *IEEE Trans. Consumer Electron.*, 39(3):455-460, Jun. 1993.

[70] M. Wu and B. Liu. Multimedia Data Hiding. Springer-Verlag Publisher, 2002.

[71] M. Wu and B. Liu. Data Hiding in Image and Video: Part-I – Fundamental Issues and Solutions. *IEEE Trans. Image Processing*, 12(6):685-695, June 2003.

[72] M. Wu, H. Yu and A. Gelman. Multi-level Data Hiding for Digital Image and Video. *SPIE Conf. Multimedia Systems and Applications*, pages 10-21, Sep. 1999.

[73] P. Yin, B. Liu and H. Yu. Error Concealment Using Data Hiding. *Proc. IEEE Int'l Conf. Acoustic, Speech and Signal Processing*, pages 1453-1456, May 2001.

[74] P. Yin, M. Wu and B. Liu. A Robust Error Resilient Approach For MPEG Video Transmission Over Internet. *SPIE Conf. Visual Comm. & Image Processing*, pages 103-111, Jan. 2002.

[75] B. Yu and K. Nahrstedt. A Scalable Overlay Video Mixing Service Model. *ACM Multimedia*, pages 646-647, Nov. 2003.

[76] W. Zeng and B. Liu. Geometric-Structure-Based Error Concealment with Novel Applications in Block-Based Low-Bit-Rate Coding. *IEEE Trans. on Circuits and Systems for Video Technology*, 9(4):648-665, Jun. 1999.

[77] X.M. Zhang, A. Vetro, Y.Q. Shi and H. Sun. Constant Quality Constrained Rate Allocation for FGS-coded Videos. *IEEE Trans. Circuits Syst. Video Technol.*, 13(2):121-130, Feb. 2003.

[78] D. Zhang and Z. Wang. Image Information Restoration Based on Long-Range Correlation. *IEEE Trans. Circuits Syst. Video Technol.*, 12(5):331-341, May 2002.