UNIVERSITY of CALIFORNIA
Santa Barbara

Error Resilient Multiple Description Video Coding over Wireless Ad-hoc Networks

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Electrical and Computer Engineering

by

Yiting Liao

Committee in Charge:

Professor Jerry D. Gibson, Chair
Professor Kenneth Rose
Professor John J. Shynk
Professor Heather Zheng

March 2011

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

The Dissertation of Yiting Liao is approved.

Professor Kenneth Rose

Professor John J. Shynk

Professor Heather Zheng

Professor Jerry D. Gibson, Committee Chair

December 2010

Error Resilient Multiple Description Video Coding over

Wireless Ad-hoc Networks

Yiting Liao

This thesis is dedicated to:

my husband, Lei Yang, and my Dad

and in the loving memory to my Mom

Acknowledgements

I would like to thank my research advisor and committee chair, Professor Jerry D. Gibson, for his support, patience, and guidance throughout the course of this dissertation. He was very instrumental in providing the right research direction while giving me the freedom to pursue ideas that intrigued me. I deeply appreciate that he was always available for discussing research problems with me and taught me how to do research. He also gave me plenty of opportunities to present my work at international conferences and helped me improve my public speaking skills. I was very fortunate to have him as my advisor.

I would also like to thank my committee members, Prof. Rose, Prof. Shynk, and Prof. Zheng for serving on my committee, and for their valuable comments and suggestions to improve the dissertation.

I am thankful to my fellow researchers in the ViVoNets Lab. Jing was the first person I talked to when I first came to the lab. She gave me a lot of help during the past four years and I always admired her ability to handle her work and personal life so well. I am also thankful to Niranjan for his help and company during my first summer in the lab. Special thanks to Malavika, it was always fun and helpful to discuss with you. Thanks to all my labmates: Steve, Pravin, and Ying-Yi, for making ViVoNets lab a fun place to work.

Most of all, I would like to thank my family for their love and support at all times. I am very grateful to my husband, Lei Yang, who has always been there for me. I must also thank my dad, for his unconditional love, support and encouragement. It saddens me that my mom will never see this thesis, as she

passed away in 2003. Without her love, I will not be the person I am today. She will always be in my heart.

Curriculum Vitæ

Yiting Liao

Education

12/2010	Doctorate of Philosophy, Electrical and Computer Engineering University of California, Santa Barbara
07/2006	Master of Science, Electronic Engineering Tsinghua University, China
07/2004	Bachelor of Science, Electronic Engineering Tsinghua University, China

Professional Experience

04/2007 - Present	Graduate Research Assistant, UCSB
06/2009 - 09/2009	Graduate Intern, Dolby Laboratories, Burbank, CA
01/2007 - 04/2007	Teaching Assistant, Department of Electrical and Computer Engineering, UCSB
07/2005 - 05/2006	Graduate Intern, STMicroelectronics China R&D Center, China
02/2004 - 06/2006	Research Assistant, Communication and Information System Laboratory, Tsinghua University

Publications

Y. Liao and J. D. Gibson, "Routing-aware multiple description video coding over mobile ad-hoc networks", *IEEE Transactions on Multimedia*, to appear, Feb. 2011.

Y. Liao and J. D. Gibson, "Routing-aware multiple description coding with multipath transport for video delivered over mobile ad-hoc networks", *the First IEEE Workshop on Multimedia Communications & Services*, to appear, Dec. 2010.

Y. Liao and J. D. Gibson, "Video communications over wireless ad-hoc networks using source coding diversity and multiple paths", Book Chapter, *Theory and Applications of Ad Hoc Networks*, to appear, Dec. 2010.

Y. Liao and J. D. Gibson, "Frame corruption estimation from route messages for video coding over mobile ad-hoc networks", *Proceedings of the 44th Annual Asilomar Conference on Signals, Systems, and Computers*, Nov. 2010.

Y. Liao, A. Leontaris, and A. M. Tourapis, "A low complexity architecture for video coding with overlapped block motion compensation", *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP)*, Sep. 2010.

Y. Liao and J. D. Gibson, "Routing-aware multiple description video coding over wireless ad-hoc networks using multiple paths", *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP)*, Sep. 2010.

Y. Liao and J. D. Gibson, "Rate-distortion based mode selection for video coding over wireless networks with burst losses", *Proceedings of the 17th International Packet Video Workshop (PV)*, May 2009.

Y. Liao and J. D. Gibson, "Enhanced error resilience of video communications for burst losses using an extended ROPE algorithm", *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2009.

Y. Liao and J. D. Gibson, "Refined error concealment for multiple state video coding over ad-hoc networks", *Proceedings of the 42nd Asilomar Conference on Signals, Systems and Computers*, Oct. 2008.

Abstract

Error Resilient Multiple Description Video Coding over
Wireless Ad-hoc Networks

by

Yiting Liao

Providing reliable video communications over wireless ad-hoc networks is becoming increasingly important as these networks become widely deployed in military, homeland defense, and disaster recovery applications. However, wireless ad-hoc networks impose great challenges to support such applications due to the highly dynamic network topology and the unreliable wireless channels. Given the error-prone nature of the wireless ad-hoc networks and the vulnerability of compressed video to packet losses, it is critical to enhance error resilience of video transmission over such lossy networks.

In this dissertation, we investigate approaches to enhance error robustness of delivered video based on multiple description coding (MDC) and a multipath transport (MPT) framework. We first investigate error concealment techniques at the decoder to improve the reconstructed video with transmission errors. The main idea is to utilize the redundancy among multiple video descriptions to provide better concealment for intra and inter frames on a macroblock (MB) basis. Since the concealment methods may cause error propagation among descriptions, we further introduce a rate-distortion optimized (RDO) mode selection

method at the encoder to enhance resilience, which is motivated by the recursive optimal per-pixel estimate (ROPE) approach. This method estimates the end-to-end distortion for MDC, considering the network conditions and multiple state recovery, and uses the estimated distortion to select the optimal coding mode and thus reduces error propagation due to packet losses. The RDO mode selection method for MDC requires knowledge of network conditions, which is time-varying and not easy to obtain in wireless ad-hoc networks.

We observe that routing messages available in the standard routing protocols indicate route changes and link failures of the network and can be used to track potential packet losses. Thus, we propose an approach that estimates the packet losses based on the routing messages and selects reference frames at the encoder accordingly. We establish a model to estimate the packet loss probability of each packet based on the routing messages and network parameters, without an additional feedback channel or extra overhead. By identifying the corrupted transmitted frames and avoiding using them as reference in multiple description coding, we reduce error propagation and improve the delivered video quality.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In recent years, there has been a growing interest in video communications over mobile wireless networks with the advances of wireless networking and video coding technologies. For example, the deployment of ad-hoc networks in military, homeland defense, and disaster recovery applications has stimulated the rapid growth of wireless multimedia applications such as video streaming to mobile devices and videoconferencing. There are also emerging applications that utilize wireless video transmissions to provide multimedia services in enterprise and home networks, remote health care and rural area networks. According to Cisco Mobile Data Traffic Forecast [2], wireless video transmissions will become the dominating part of total wireless data traffic. By 2014, almost 66% of world's mobile data traffic will be video.

Despite the increasing demand of wireless video applications, providing reliable video communications over wireless ad-hoc networks faces significant challenges. The challenges arise from both the error-prone nature of the wireless networks and the vulnerability of compressed video to packet losses. While offer-

ing more flexible connections than wired networks, mobile ad-hoc networks have several features that make video transmissions challenging. First, node mobility and the lack of infrastructure in the network can lead to frequent topological changes, and consequently link failures and route changes. Second, the wireless link quality is affected by channel fading and shadowing, as well as noise. Such time-varying and location-dependent channel conditions can translate into random and burst packet losses in the network. Finally, wireless transmissions suffer from interference from nearby transmissions. As deployments continue to grow, the devices may suffer from excessive interference, which can also lead to significant degradation in video performance.

The most widely used video coding technique called "block-based motion estimation and compensation" makes the compressed video very vulnerable to packet losses. In motion-compensated prediction, the current frame is predicted from the previous encoded frames. Therefore, any errors in a video frame may propagate to subsequent frames and cause severe distortion in the reconstructed video.

Given the challenges above, it is important to design error resilience mechanisms for reliable video transmission over wireless ad-hoc networks. The existing solutions can be categorized into three groups: (1) Forward techniques that add a controlled amount of redundancy to the transmitted video bitstream, such as forward error correction [3], intra/inter mode selection [1, 4], and multiple description coding [5]; (2) concealment techniques that attempt to conceal the effects of errors at the decoder, such as spatial interpolation [6], recovery of motion vectors [7], and hybrid concealment algorithms [8]; (3) interactive error

control that requires the interaction between the encoder and the decoder, such as automatic repeat request (ARQ) [9], and reference picture selection [10, 11].

Among these error resilient techniques, multiple description coding (MDC) with path diversity has been shown to be a promising technique for video transmission over lossy networks [5]. MDC generates multiple descriptions with equal importance while each description can reconstruct the source with acceptable quality. In wireless networks, packets suffer from bursty losses due to node mobility and topology changes. With multiple paths, MDC is robust to bursty losses because acceptable video quality can be maintained as long as not all descriptions are lost simultaneously. Unlike scalable video coding methods which require successful reception of the base layer, MDC treats each description equally and can construct video from any received description. In addition, MDC with path diversity can be used to distribute and balance the load in the network, thus avoiding network congestion. In this dissertation, we design enhanced error resilient video coding techniques based on a multiple description coding/multipath transport framework.

1.1 Objective and Major Contributions

The objective of this research is to support reliable video communications over wireless networks using multiple description video coding with path diversity. Our work is built on a multiple description coding/multipath transport framework. We propose three techniques to improve the delivered video quality over the lossy networks. Our objective and contributions are illustrated in

Fig. 1.1. At the encoder, we develop a forward technique that estimates the

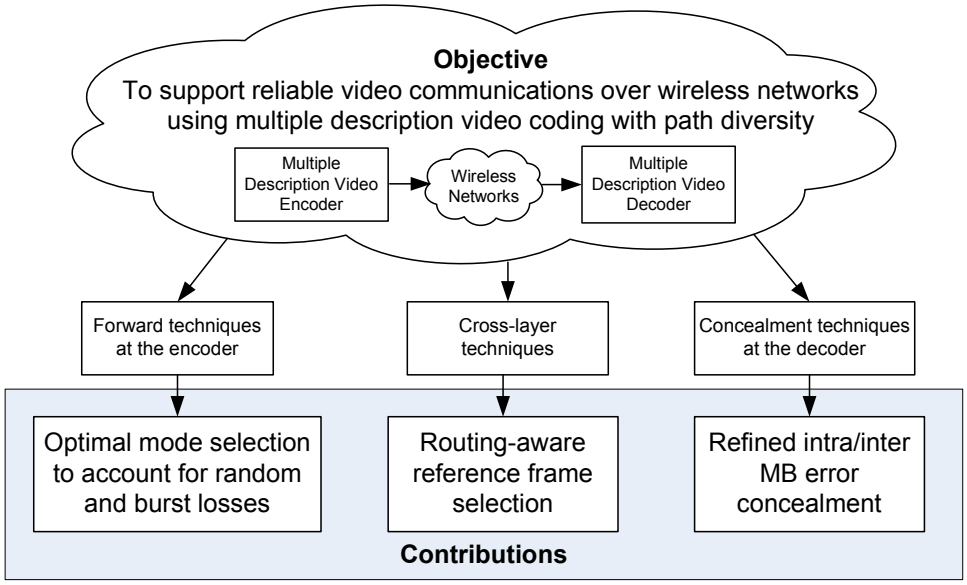


Figure 1.1: Dissertation Overview: Objective and Contributions

end-to-end distortion and selects the optimal coding mode for MDC. We also propose a cross-layer technique that utilizes routing messages to estimate the packet loss probability of each packet and adaptively selects reference frames to reduce error propagation. At the decoder, we apply refined error concealment methods for MDC to better reconstruct the corrupted video. Next, we discuss our major contributions in details.

First, we propose MB-based intra and inter error concealment methods for the multiple description video decoder. Since a certain amount of redundancy exists among multiple video descriptions, it is important to utilize their correlation to conceal the errors. For intra coded macroblocks (MBs), we use temporal correlated MBs in the adjacent intra frame in the other description as candidates for concealment and the candidate that minimizes the side match distortion is

used to conceal the lost intra MB. For inter coded MBs, additional decoded frames from the other description are added to the reference list to assist the motion compensated concealment. Compared to the original error concealment method in multiple state video coding (MSVC), our proposed method provides better reconstruction in a MB level and achieves performance gains across a wide range of random and burst loss rates.

To enhance the error robustness of the video bitstream, we adopt a rate-distortion optimized mode selection method at the encoder that accounts for random and burst losses in the networks. MDC is considered as an effective method to combat burst losses in lossy networks because even if a whole description is lost, the remaining description can still be decoded with acceptable video quality. However, when MDC experiences random packet loss, the distortion due to random loss not only propagates to subsequent frames in the same description, but may also affect frames in the other description because of the multiple state recovery at the decoder. By estimating the end-to-end distortion for MDC at the encoder, and selecting coding mode under rate-distortion optimized constraints, we improve the robustness of the video bitstream to both random and burst losses.

Finally, we discover that standard routing messages in the wireless ad-hoc networks provide some link failure information and can be used to track the packet loss without introducing extra overhead. We build a model to characterize the relationship between routing messages and packet loss probability of transmitted packets. The proposed model has been shown to effectively estimate the real-time packet loss information. Furthermore, we adaptively select

reference frames based on the estimation to reduce the error propagation caused by the corrupted frames.

1.2 Dissertation Organization

This dissertation is organized as follows. Chapter 2 provides basic background on video compression and transmission. We first present the general architecture of a video codec. Then we review various multiple description video coding methods and introduce MSVC, which has been employed in this dissertation. We also introduce the video quality metrics used to evaluate the video quality.

In Chapter 3, we study error concealment methods for MSVC. Exploring correlation across different descriptions in MSVC is beneficial to improve video quality over lossy networks. We present our MB-based refined error concealment method for MSVC and show its effectiveness under a variety of packet loss patterns.

In Chapter 4, we present our rate-distortion optimized mode selection method that accounts for both random and burst losses. This method is applied to both single description coding and multiple description coding to enhance the error robustness of the transmitted video.

In Chapter 5, we describe a cross-layer framework to support video transmission over mobile ad-hoc networks with multiple path transport. The routing messages available in the standard routing protocols are used to estimate the

potential packet losses in the networks and then the encoder dynamically selects reference frames based on the estimates to alleviate error propagation.

Chapter 6 concludes this dissertation and discusses some future work.

Chapter 2

Background

In this chapter, we introduce some basic concepts that are closely related to this dissertation. We first introduce the general architecture of a video codec, and then discuss the multiple state video coding approach that we adopt to most of our work. We finally introduce the video quality measures we utilize to evaluate the performance of our approaches.

2.1 General Architecture of Video Codecs

For the past two decades, video coding techniques and standards have been an important research and development topic in both industry and academia. The majority of video standards are developed by two groups: Video Coding Experts Group (VCEG) of the International Telecommunication Union Telecommunication Standardization Sector (ITU-T), and the Moving Picture Experts Group (MPEG) of the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC). Figure 2.1 shows the

chronological development and applications of video standards published by VCEG and MPEG. Currently, H.264/AVC is the most recent and advanced video standard widely adopted in the markets.

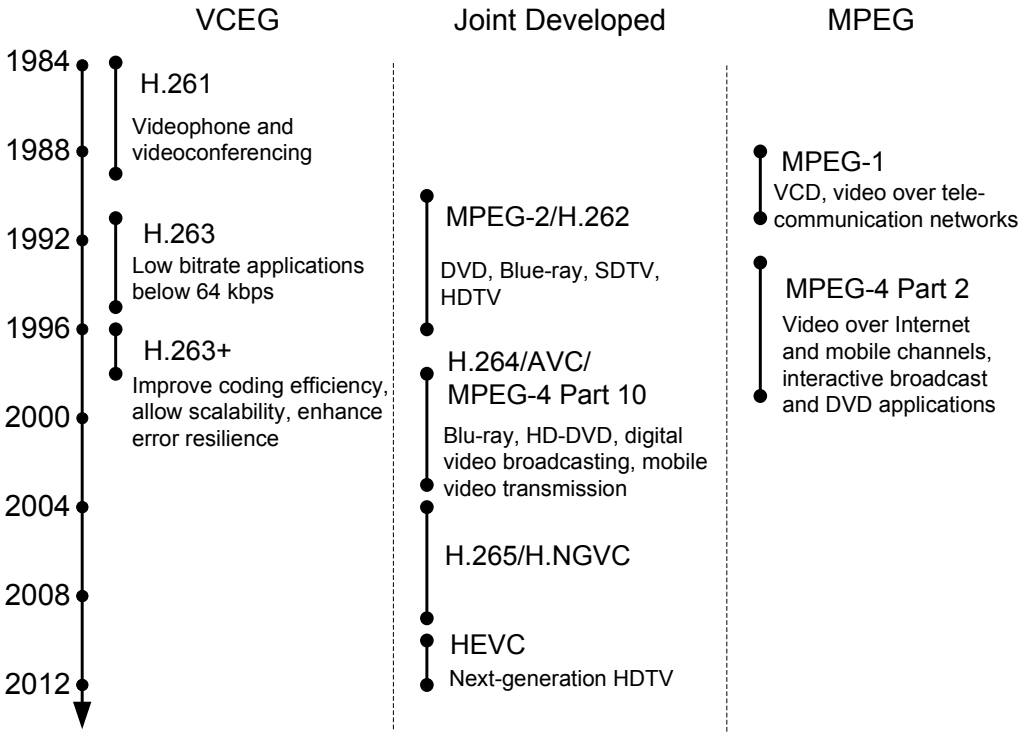


Figure 2.1: Chronological Development of Video Standards Published by VCEG and MPEG

Even though different video coding techniques have been adopted in different video standards, most video codecs follow a general architecture as shown in Fig. 2.2. The video signals are passed through a motion compensated prediction block or an intra prediction block to remove their temporal or spatial redundancy. Then the transform block converts residues into transform coefficients. The quantization block removes the insignificant coefficients and outputs a set of quantized transform coefficients. Finally the entropy coder

compresses the information including motion vectors, quantized residual coefficients, header information to generate the coded bit stream. At the decoder, the inverse operations are performed to reconstruct the video for display.

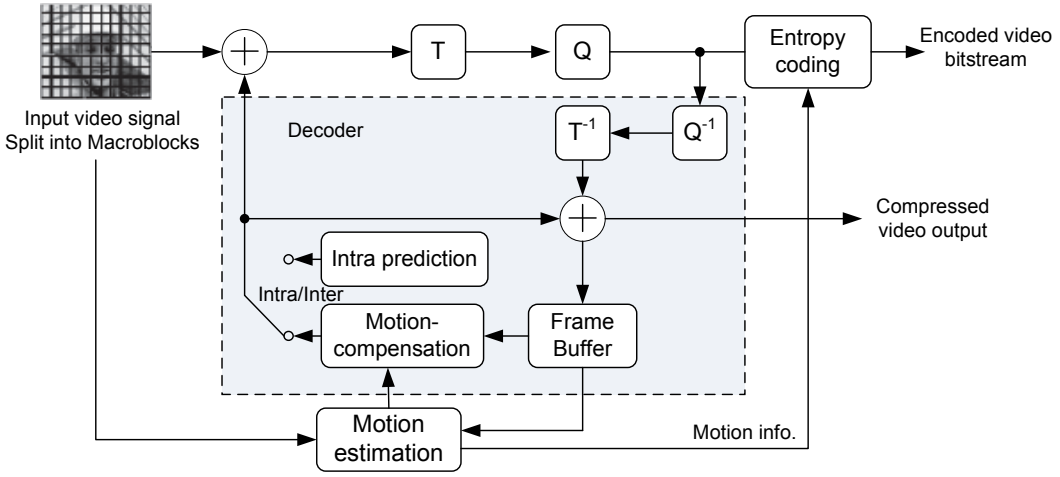


Figure 2.2: General Architecture of Video Codecs

The core of the encoder is the "block-based motion estimation and compensation", which reduces the temporal redundancy by subtracting the predicted block from the current block. During the motion estimation (ME) process, the encoder estimates the motion between the reference frame and current frame, and finds a block in the reference frame that best matches the current block. The encoder then generates the residue block by subtracting the predicted block in the reference frame from the current block. This process is called motion compensation (MC). Block-based motion estimation and compensation is the most widely used video coding technique because it can effectively remove the temporal redundancy in a video sequence and it fits well with the block-based transform techniques. However, there are several disadvantages. The motion

estimation and compensation module contributes to most of the complexity at the encoder. Real world objects do not typically have clean edges with rectangular boundaries, and many types of motion, such as zooming, rotation, and warping, are difficult to compensate using block-based methods. Furthermore, motion compensated prediction makes the video bitstream very vulnerable to errors. As described, a video frame is usually predicted from a reference frame. If the reference frame contains some errors, these errors may also occur in the frames that are predicted from the corrupted reference frame. In other words, errors in one frame may lead to errors in the subsequent frames and cause severe damage to the reconstructed video. This is called error propagation. Most of the work in this dissertation focuses on alleviating error propagation and improving delivered video quality.

2.2 Multiple Description Video Coding

MDC is an effective approach to enhance the error resilience of video transmission over lossy networks. The general idea is to encode the video sequence into several descriptions with equal importance. Each description can be decoded independently or combined with other descriptions for reconstruction. In general, the reconstructed video achieves better video quality when more descriptions are received.

Many MDC algorithms have been proposed [12] and they can be divided into three categories: subsampling algorithms in the temporal [13], spatial [14] or frequency domain [15], multiple description quantization algorithms [16, 17],

and multiple description transform coding [18]. Wang et al. provide a good review for MDC algorithms [5].

Since subsampling methods are easy to implement and compatible with different video standards, they have been the most commonly investigated MDC algorithms. These methods generally work in the spatial, temporal, or frequency domain to generate multiple descriptions, and the corresponding correlation is used to recover a lost description. One of the most popular MDC methods is multiple state video coding (MSVC) [13], which temporally downsamples the video sequence and uses the correlation between adjacent frames in two descriptions to recover from frame loss. In MSVC, the system includes a multiple state video encoder/decoder and a path diversity transmission system as shown in Fig. 2.3.

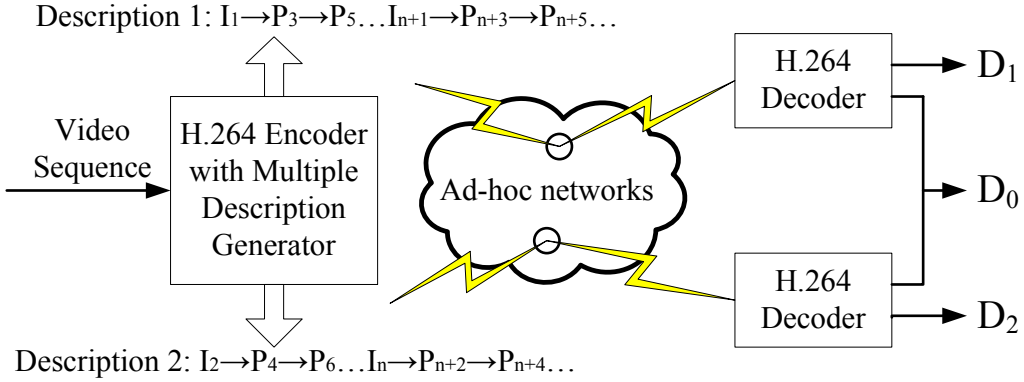


Figure 2.3: MSVC System Architecture

At the encoder, the video sequence is first temporally down-sampled into two sub-sequences, i.e. odd frames in the original sequence are extracted as one sub-sequence and even frames as the other. The two sub-sequences are encoded separately using a H.264 video encoder [19] and transmitted over the network

in two different paths. At the decoder, they are decoded and interleaved to get the reconstructed video sequence. Most of the work in this dissertation is built on the MSVC method.

2.3 Video Quality Measurement

Video quality measurement plays an important role in designing, evaluating, and optimizing video compression and transmission techniques. A reliable way to evaluate video quality is subjective measurements, in which a group of subjects watch the reconstructed video and designate their opinions on the perceived quality. Despite their reliability, subjective measurements are expensive and time consuming, while objective measures can be cheaper and faster alternatives. Average peak signal-to-noise ratio (PSNR) is the most widely used objective video quality measure. PSNR represents the mean squared error (MSE) of the distorted video and is defined by

$$PSNR = 10 \cdot \log_{10} \frac{255^2}{MSE} \tag{2.1}$$

where MSE is the mean square error between the original pixel and the distorted pixel. PSNR has been used by the majority of video coding standards including H.264. However, video sequences with close average PSNR across all frames and all realizations may reveal very different video quality for different users. Therefore, we also utilize $\text{PSNR}_{r,f}$ proposed by Hu et al. [20] to evaluate the video quality for multiple channel uses. $\text{PSNR}_{r,f}$ helps to capture various attributes of the video which are averaged out and invisible in the conventional average PSNR measure. $\text{PSNR}_{r,f}$ is defined as the PSNR achieved by $f\%$ of

the frames for $r\%$ of realizations, which shows the video quality guaranteed for $r\%$ of realizations among $f\%$ frames. The definition of $\text{PSNR}_{r,f}$ can be written as

$$PSNR_{r,f} = \arg_x P_{real}(P_{frame}(PSNR > x) \geq f) \geq r) \tag{2.2}$$

Here, $P_{frame}(PSNR > x)$ is the percentage of frames that have PSNR higher than x in a realization and $P_{real}(\Omega)$ is the percentage of realizations that satisfy the condition Ω. For example, $\text{PSNR}_{r=80\%,f=90\%} = 35$ dB means that there are 80% of the realizations having 90% of their frames with PSNR higher than 35 dB. We use $\text{PSNR}_{r,f}$ as a multiuser perceptual video quality indicator because of two reasons. First, $\text{PSNR}_{r,f}$ captures the lowest PSNR achieved by $f\%$ of the frames in each realization, which can be used to measure the perceptual video quality of a single realization due to two observations in video quality assessment [21]: (1) The bad-quality frames in a video dominate users' experience with the video; (2) For PSNRs higher than a certain threshold, increasing PSNR does not help to enhance the perceptual video quality. Unlike average PSNR that treats every frame equally, $\text{PSNR}_{r,f}$ captures the performance loss due to damaged frames in a video sequence ($f\%$). Second, due to the time-varying network conditions, multiple users or a user who accesses the network multiple times may have different experiences. $\text{PSNR}_{r,f}$ can capture the performance experienced by a user in multiple uses ($r\%$), or alternatively, it indicates the percentage of video users (r%) that experience a specific video quality.

Chapter 3

Refined Error Concealment for MSVC

Multiple description coding (MDC) with path diversity is often used to maintain connectivity in wireless ad-hoc networks, but the challenge is to match the chosen MDC method with an effective error concealment scheme. In this chapter, we develop a macroblock (MB) based refined error concealment method for MSVC by exploring the correlations across different descriptions. The refined intra MB concealment provides better concealment for MBs in intra frames by using temporal correlation between adjacent intra frames in two descriptions. The refined inter MB concealment achieves improvement from the additional reference frames used for motion-compensated concealment. We show that the proposed refined error concealment method can effectively enhance the error resilience of MSVC over error-prone wireless networks.

This chapter is organized as follows. In Section 3.1, we survey related work on error concealment techniques for video coding. In Section 3.2 and Section 3.3,

we propose MB-based error concealment methods for intra and inter frames in MSVC respectively. In Section 3.4, the performance of proposed method is evaluated over a wireless network with both random and burst losses.

3.1 Existing Error Concealment Techniques for Video Coding

Error concealment techniques, which have been well developed for decades [22], make use of the spatial and temporal correlation between video pixel values to recover a corrupted video stream with random channel errors.

Some error concealment techniques explore the spatial redundancy of video sequences for spatial domain or transform domain reconstruction. Aign and Fazel proposed to interpolate lost pixel values from the boundary pixels of the four neighboring MBs [23]. In addition, [6, 24–28] used different algorithms to detect the edges within the lost MBs and directionally interpolate the lost pixels along the edges. A more complex approach called coarse-to-fine block replenishment (CFBR) [29] performed the interpolation by first recovering the smooth large-scale patterns, then the large-scale structures, and finally the local edges in the lost MB. In addition to reconstruction in the spatial domain, a number of papers address the transform coefficient recovery problem by interpolating the lost coefficient from corresponding coefficients in the neighboring MBs [30, 31], performing optimization based on a smoothness constraint [32, 33], using the fuzzy logic approach to recover the high-frequency components [34], or using an iterative procedure called "projections onto convex sets" (POCS) [35].

Other concealment techniques exploit temporal redundancy to estimate the lost motion information and replace the lost MB with the motion-compensated MB from one of the previous frames. Numerous approaches have been studied to recover the lost motion vectors (MVs). Haskell and Messerschmitt discussed the use of zero MV, the MV of the co-located MB in the previous frame, and the average or median MV of the spatially adjacent MBs to recover the lost MB [36]. The boundary matching algorithm (BMA) [37] is proposed to select the best MV among a set of candidate MVs. [7, 27, 38, 39] presented different block matching techniques that estimate the MV based on the set of surrounding MBs of the lost MB. Salama et al. modeled the motion field as a Markov random field (MRF) and found the maximum a posteriori (MAP) estimate of the lost MV given its neighboring MVs [40]. This method is further improved by using an adaptive Huber function in an MRF model [41].

More recently, hybrid algorithms have been proposed to obtain better recovery. They are effective but generally introduce more complexity. Shirani et al. first obtained initial estimates of the missing MB by motion compensation or spatial interpolation and then used a MAP estimator to refine the initial estimates [42]. Atzori et al. proposed a concealment method which replaces the lost MB using BMA and applies a mesh-based warping (MBW) to reduce the artifacts [43]. In [8], the lost MV is first estimated by a spatio-temporal BMA algorithm, and a partial differential equation (PDE) based algorithm is used to refine the reconstruction.

These error concealment techniques can be exploited to fill in lost data, however, the effectiveness of these traditional methods is constrained by the fact

that the information available across descriptions are not exploited. Therefore, some studies propose error concealment methods targeted for different MDC methods to better utilize the information available in descriptions. Lee and Altunbasak [44] adopted a MAP estimation approach to conceal the corrupted description in multiple description transform coding [18] and Wang et al. [45] proposed an error concealment method for a three-loop slice group MDC approach [46]. In [13, 47, 48], different concealment methods are proposed to recover the lost frame in MSVC. However, these recovery approaches are designed to recover the loss of an entire frame, while a video bitstream transmitted over wireless networks may suffer random packet loss that causes only some MB losses. In the next two sections, we introduce the MB-based error concealment methods for intra and inter MBs in MSVC respectively.

3.2 Refined Intra MB Concealment for MSVC

In H.264, the lost MB in an intra frame is concealed spatially based on weighted pixel interpolation [49]. As shown in Fig 3.1, each pixel in the lost MB is estimated from the weighted sum of the boundary pixels in the adjacent MBs, where the weight is the inverse distance between the pixel to be concealed and the boundary pixel. In other words, the lost pixel can be calculated by

$$Y(x,y) = \frac{\sum_{i=1}^{4} Y_i(16-d_i)}{\sum_{i=1}^{4}(16-d_i)} \tag{3.1}$$

where d_i is the distance between the concealed pixel and the boundary pixel in the neighboring MB, and Y_i is the boundary pixel value as shown in Fig. 3.2.

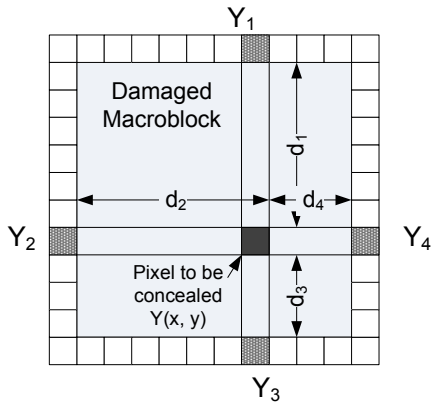


Figure 3.1: Intra MB Concealment in H.264

Only the correctly received neighboring MBs are used for the concealment unless less than two correctly received MBs are available. In that case, the neighboring concealed MBs are also used for the interpolation. For SDC, each group of picture (GOP) only contains one intra frame. In order to stop the error propagation from the previous GOP, the lost MBs in an intra frame in SDC is only concealed spatially. For MSVC, each description has an intra frame in every GOP and the two intra frames are consecutive as shown in Fig. 2.3. Therefore, we can apply both temporal and spatial concealment for the lost MBs in the two consecutive intra frames for MSVC.

The process to conceal lost MBs in the two consecutive intra frames is shown in Fig. 3.2. First, the correctly received MBs in the two intra frames are decoded. Then for the MBs that are lost at the same spatial position in both intra frames, the weighted pixel interpolation method shown in Fig. 3.1 is applied for concealment. For other lost MBs, we copy the MBs in the corresponding position in the other intra frame and calculate the side match distortion [37] based on the correctly received neighbor MBs. As shown in Fig.

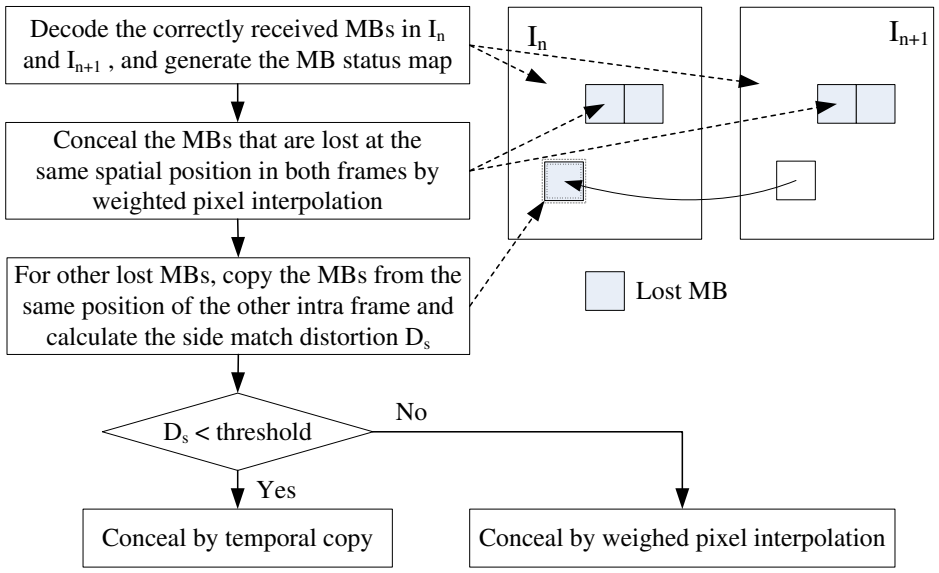


Figure 3.2: Error Concealment for MBs in an Intra Frame for MSVC

3.3, the side match distortion D_{SM} is the sum of absolute luminance differences between the concealed MB and neighboring MBs at the boundary,

$$\begin{aligned} D_{SM} &= \sum_{i=0}^{15} |Y_{x_0+i,y_0} - Y_{x_0+i,y_0-1}| + \sum_{i=0}^{15} |Y_{x_0,y_0+i} - Y_{x_0-1,y_0+i}| \\ &\quad + \sum_{i=0}^{15} |Y_{x_0+i,y_0+15} - Y_{x_0+i,y_0+16}| + \sum_{i=0}^{15} |Y_{x_0+15,y_0+i} - Y_{x_0+16,y_0+i}| \end{aligned} \quad (3.2)$$

We compare the side match distortion to a pre-defined threshold. If the side match distortion is smaller than the threshold, the temporal copy concealment is applied to conceal the lost MB. If not, the weighted pixel interpolation is used to conceal the lost MB.

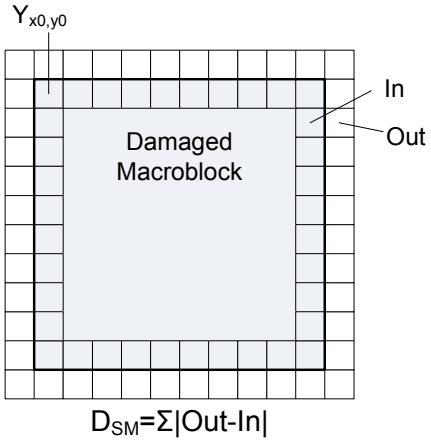
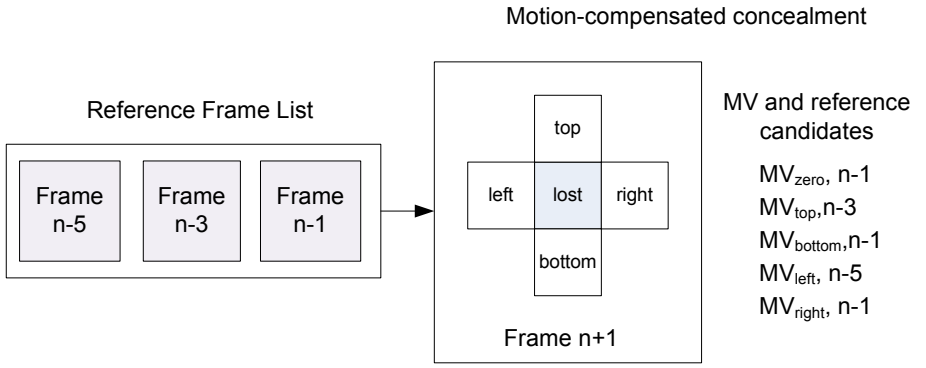


Figure 3.3: Side Match Distortion for a Damaged MB

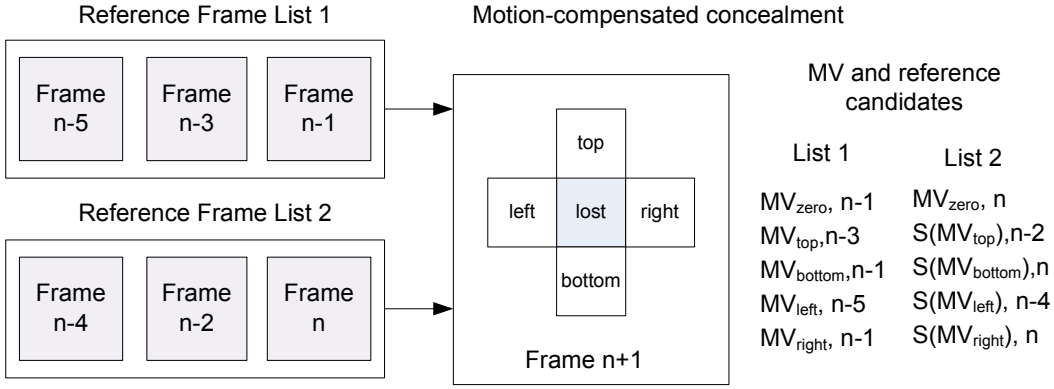
3.3 Refined Inter MB concealment for MSVC

In H.264 reference software, the lost MB in the inter frame is concealed by estimating the lost motion vector from the neighbor MBs and applying motion-compensated prediction [50]. When an inter MB is lost, the motion vector of the missing MB is predicted from one of the neighbor MBs or zero motion vector as shown in Fig. 3.4(a). The motion vector that has the minimum side match distortion is used for motion-compensated concealment. The reference frames used to conceal the lost MB are the same as the reference frames for correctly received MBs.

Since the even and odd frames are encoded independently in MSVC, the correlation between the reference frame and the current frame is reduced. If we only use the frames in the same description as the reference frames to conceal the lost MB, it may not perform as well as using the reference frames from the other description for concealment. Therefore, we propose to explore the



(a) Inter MB Concealment in H.264



(b) Refined Inter MB Concealment for MSVC

Figure 3.4: Inter MB Concealment Methods

information from both descriptions in MSVC to enhance the inter error concealment; that is, we use two reference frame lists from each description for the motion-compensated concealment. The reference list that results in better side match distortion is used as the reference to recover the lost MBs.

In order to perform motion compensated concealment for inter MBs, we need to estimate the lost motion vector and the corresponding reference frame. As shown in Fig. 3.4(b), instead of using only frames in the same description as

reference frames, we add reference frame list 2 from the other description. Then we use MVs from four neighboring MBs and zero MV (shown in Fig. 3.4(b)) as MV candidates, and apply motion compensated concealment by using the corresponding reference frame from reference frame list 1. When reference frame list 2 is used for concealment, we need to scale these MV candidates accordingly because the estimated motion vector corresponds to the reference frame in list 1. Assume one of the MV candidates is $MV_{candidate}$ with reference frame n_1, and its corresponding reference frame in reference frame list 2 is n_2, then the scaled MV for reference frame list 2 can be calculated by

$$S(MV_{candidate}) = \frac{n_c - n_2}{n_c - n_1} MV_{candidate} \tag{3.3}$$

where n_c is the current frame number. Similarly, motion-compensated concealment is applied based on the scaled MVs and the reference frames in the reference frame list 2. Finally, we choose the estimated motion vector and reference list that minimizes the side match distortion to conceal the lost inter MB.

3.4 Performance Evaluation in Wireless Ad-hoc Networks

In this section, we first present a packet loss model to simulate the random and burst packet losses in wireless networks. Then we introduce the parameters and methods to encode the video for comparison. Finally, the performance

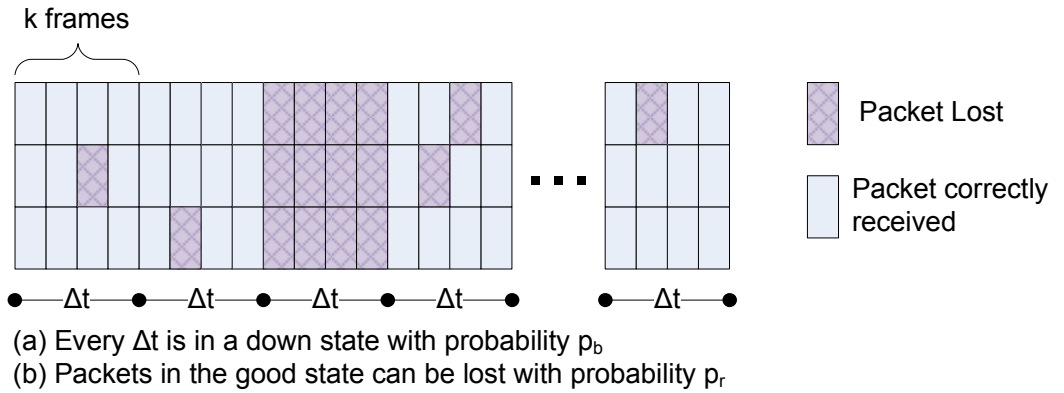


Figure 3.5: Packet Loss Model for Wireless Ad-hoc Networks

of our proposed method under different network conditions for different video sequences is evaluated.

3.4.1 Packet Loss Model

In wireless networks, packet loss may occur due to numerous reasons, including link/node failures, route changes, and bit errors. These factors can cause both random packet loss and burst losses over the network. To investigate the video communications over such lossy networks, we introduce a packet loss model that captures packet loss features in the network. As shown in Fig. 3.5, this model considers both random packet loss and burst losses during transmission and can be used to generate different loss patterns over the wireless network.

In this model, time is divided into Δt intervals and k frames are transmitted during an interval. Each interval may be either in a good state with probability $(1 - p_b)$ or in a down state with probability p_b, which is independent and iden-

tically distributed. The packets transmitted in a down state are all lost while the packets transmitted in the good state may suffer from a random packet loss. Therefore, the packet loss model can be determined by three parameters: the burst loss rate p_b, the burst length k (frames), and the random packet loss rate p_r in a good state. The total packet loss rate p in the networks can be calculated by

$$p = p_b + (1 - p_b)p_r = p_b + p_r - p_b p_r \tag{3.4}$$

3.4.2 Compared Schemes and Simulation Settings

We implement our proposed methods by modifying H.264 reference software JM13.2. To evaluate the performance of the refined error concealment method for MSVC, we compare the following three coding schemes:

- **SDC**: The video sequence is coded into a single description and transmitted over one path over the network.
- **MSVC**: The video sequence is coded into two descriptions using MSVC and transmitted over two independent paths over the network.
- **MSVC_REC**: The video sequence is coded and transmitted the same as MSVC, while the refined error concealment method introduced in Section 3.2 and 3.3 is applied to decode the corrupted video.

We evaluate six video sequences including Carphone, Claire, Foreman, Hall-monitor, Mother-daughter, and News. Each sequence consists of 300 frames at QCIF format. The sequences are encoded as IPPP format with $GOP = 30$ at

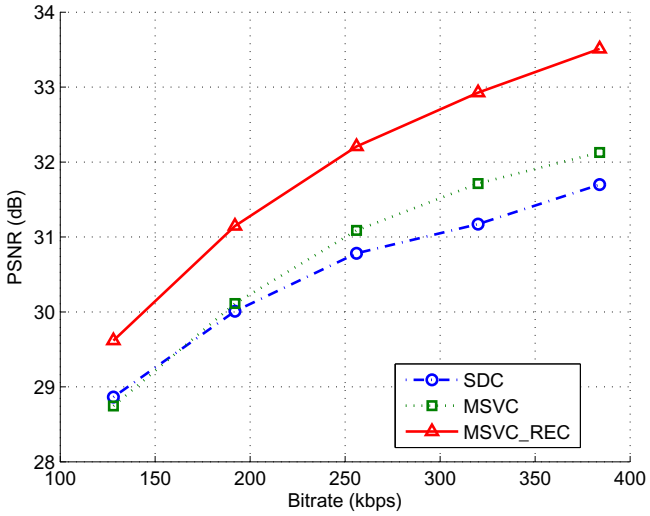
30 fps and each frame is packetized to 4 RTP packets. We examine the video performance under various bitrates (128 - 384 kbps). The video packet size is defined by

$$l = \frac{r}{8f \cdot n}(bytes/packet) \tag{3.5}$$

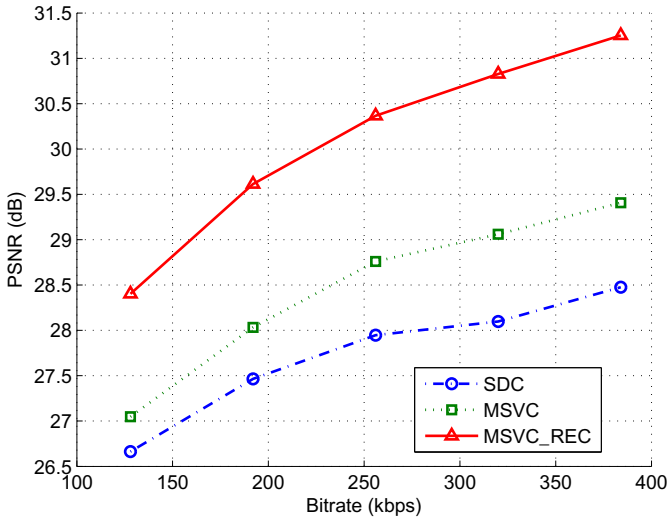
where r is the bitrate of the encoded video, f denotes the frame rate, and n represents the number of packets per frame. Based on the above settings, the video packet sizes under different bitrates are in the range of 133 - 400 bytes, which are reasonable packet sizes for wireless transmission [51]. The packet loss model in Section 3.4.1 is used to simulate the random and burst losses in wireless networks and we simulate each video sequence over 500 different realizations for each network setting.

3.4.3 Overall Performance of MSVC with Refined Error Concealment

In Fig. 3.6, we show the PSNR performance at two packet loss rates under different bitrates for Foreman sequence. Figure 3.6(a) compares SDC, MSVC, and MSVC_REC under $p_r = 2\%$, $p_b = 2\%$, and $k = 5$. Under this network condition, SDC and MSVC have similar PSNR performance at low bitrates. This is because even though the usage of multiple descriptions and path diversity enhances the robustness of MSVC, the decreased correlation between adjacent frames in each description reduces its coding efficiency. At a low packet loss rate, the gain in error resilience may not compensate for the reduction in coding efficiency, while MSVC achieves more gain at a higher packet loss rate as



(a) $p_r = 2\%, p_b = 2\%, k = 5$



(b) $p_r = 4\%, p_b = 4\%, k = 5$

Figure 3.6: Average PSNR vs Bitrate for SDC, MSVC, and MSVC_REC, Foreman Sequence at 30 fps

shown in Fig. 3.6(b). Compared to SDC and MSVC, our proposed MSVC_REC achieves consistent gains under different bitrates and network conditions. The

gains achieved by MSVC_REC are in the range of 0.8-1.8 dB at a 4% packet loss rate (Fig. 3.6(a)) and in the range of 1.4-2.8 dB at a 8% packet loss rate (Fig. 3.6(b)) for Foreman sequence. This shows that our proposed method can effectively improve the error concealment for MSVC by utilizing the information from both descriptions to conceal the packet loss on a MB basis.

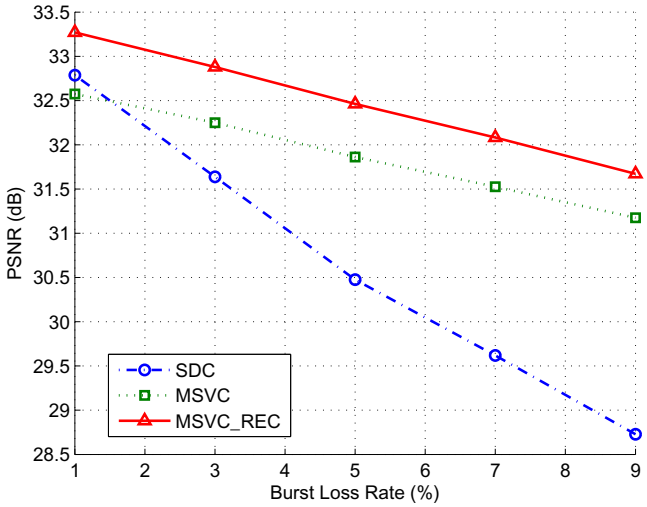
In Table 3.1, we compare SDC, MSVC, and MSVC_REC for different video sequences at 256 kbps. The results show that the average PSNR gains of MSVC_REC over SDC and MSVC are 1.52 dB and 1.58 dB for the six video sequences under an overall packet loss rate of 4%, and the gains of MSVC_REC over SDC and MSVC under an overall packet loss rate of 8% are 2.80 dB and 2.44 dB. This table illustrates that MSVC_REC outperforms the other two methods for different video sequences under different network conditions.

Table 3.1: Average PSNR (dB) for Different Video Sequences, 30 fps, 256 kbps

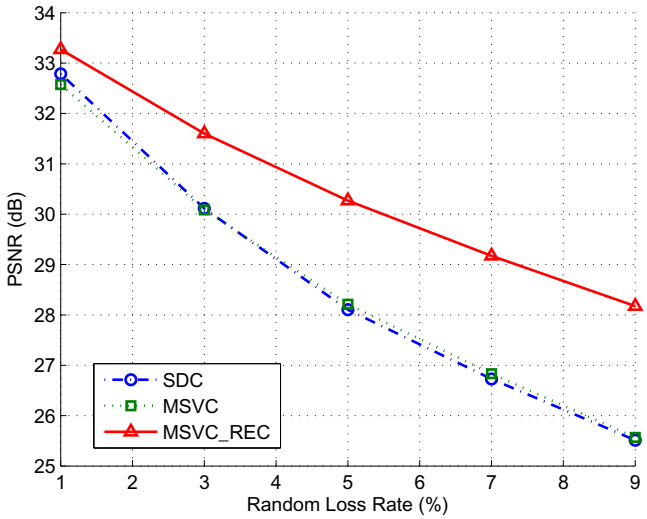
Sequence	$p_b = 2\%, p_r = 2\%, k = 5$			$p_b = 4\%, p_r = 4\%, k = 5$		
	SDC	MDC	MDC_REC	SDC	MDC	MDC_REC
Carphone	31.11	31.93	33.21	27.91	29.21	31.02
Claire	40.37	39.84	42.31	37.03	36.61	40.42
Foreman	30.78	31.09	32.21	27.95	28.76	30.37
Hall-monitor	36.68	36.68	38.46	33.36	33.61	36.52
Mother-daughter	38.11	37.69	39.20	35.64	35.42	37.89
News	36.32	35.77	37.08	32.77	33.15	35.20

3.4.4 Impacts of Random Loss Rate and Burst Loss Rate

Figure 3.7 compares PSNR performance of SDC, MSVC, and MSVC_REC under various random loss rates and burst loss rates for Foreman sequence at



(a) $p_r = 1\%, p_b = 1\% - 9\%, k = 5$



(b) $p_r = 1\% - 9\%, p_b = 1\%, k = 5$

Figure 3.7: Average PSNR vs Packet Loss Rate for SDC, MSVC, and MSVC_REC, Foreman Sequence at 30 fps, 256 kbps

256 kbps and shows that the proposed MSVC_REC effectively enhances error resilience of video under a wide range of random loss rates and burst loss rates.

In Fig. 3.7(a), the random loss rate is fixed at 1% and the PSNR performance of SDC, MSVC, and MSVC_REC under different burst loss rates is shown. We notice that the performance of SDC drops more quickly than MSVC and MSVC_REC as the burst loss rate increases, which means that SDC is more vulnerable to burst losses. That is because even if one description for MSVC is totally lost, the other description can still be correctly decoded and used to recover the lost description. Figure 3.7(a) shows that MSVC and MSVC_REC are more effective to combat burst losses than SDC. MSVC_REC has higher PSNR than MSVC of about 0.6 dB under various burst loss rates.

Figure 3.7(b) investigates the performance of SDC, MSVC, and MSVC_REC under different random loss rates with a fixed burst loss rate at 1%. MSVC leads to reduced coding efficiency due to reduced correlation between adjacent frames in each description, while providing extra error resilience. As a result, MSVC achieves comparable PSNR performance to SDC under various random packet loss rates. We see that MSVC_REC achieves up to 2.6 dB gains in PSNR and the performance gains of MSVC_REC increase as the random packet loss rate increases. This is because with the refined error concealment methods, MSVC_REC better exploits the correctly received information from both descriptions to conceal the random lost MBs.

Chapter 4

Rate-distortion Optimization for MSVC

Most error resilient techniques simply consider the average packet loss rate to enhance error robustness for video transmission. However, loss patterns, specifically burst losses, have great impact on video quality [52]. In this chapter, we propose a method that can take account of both random and burst losses to further improve the error resilience of video coding. Our method estimates the end-to-end distortion based on recursive optimal per-pixel estimate (ROPE) including both random and burst losses, and applies it for rate-distortion (RD)-based optimal mode selection. We apply our method in two cases: For single description video coding, we estimate the reconstructed pixel values for random packet loss and burst losses, and calculate the overall distortion. For multiple description video coding, we estimate the end-to-end distortion for MSVC by considering the network conditions and multiple state recovery to reduce the error propagation due to packet loss in both descriptions for MSVC. Simulation

results show that our proposed method achieves better performance than MSVC and original ROPE (only considering average packet loss rate) over wireless networks with random and burst losses.

This chapter is organized as follows. The existing work on error resilient video coding is reviewed in Section 4.1. Section 4.2 introduces the RD-based mode selection method with recursive optimal per-pixel estimation. We present our proposed method for SDC and MSVC in Section 4.3 and Section 4.4 respectively. The proposed method is compared with ROPE and MSVC under different loss patterns, and the simulation results are presented in Section 4.5. Section 4.6 discusses the complexity of the proposal methods and estimates the performance with mismatch in the network conditions.

4.1 Error Resilient Video Coding

In wireless networks, video transmission may suffer from packet loss due to link errors, node failures, route changes, interference and fading in the wireless channel. Packet loss can seriously degrade the received video quality, especially due to the propagated errors in the motion-compensated prediction loop. Therefore, it is challenging to provide error resilient video coding for reliable video communications over such lossy networks. A number of techniques have been proposed to increase the robustness of video communications to packet loss, such as intra/inter mode selection [1, 4, 53–58], reference picture selection [10, 11], and multiple description video coding [5].

Intra coding is an important technique for mitigating error propagation due to packet loss and makes the video stream more robust to errors. However, using more number of intra-coded MBs can greatly reduce the coding efficiency since an intra-coded MB generally requires more bits than an inter-coded MB. Therefore, to select the optimal intra/inter mode that can achieve the best tradeoff between error robustness and coding efficiency has become a widely addressed problem. There are some simple intra updating methods such as refreshing contiguous intra blocks periodically [53], or intra-coding blocks randomly [55].

A more advanced category of intra refresh algorithms estimates the end-to-end distortion due to both compression and packet loss, and incorporates mode selection with rate-distortion (RD) optimization [1, 4, 54–58]. An early work of RD-based mode selection method is proposed in [54], in which the distortion is roughly estimated. In [55], the encoder considers the effects of error concealment and encodes the area that is severely affected by packet loss in the intra mode. However, the error propagation beyond one frame is ignored during the estimation procedure. In [56], the authors further incorporate the distortion due to error concealment of a current block with the distortion due to error propagation from concealed blocks to optimize mode selection. One drawback of the methods proposed in [54], [55], and [56] is that the estimated distortion at the encoder is not very accurate.

A more precise approach to estimate the end-to-end distortion is proposed by [57]. The authors generate K copies of the channel behavior at the encoder and calculate the decoder reconstruction to estimate the expected end-to-end distortion. This approach can accurately estimate the distortion if K is large

enough. However, it has extremely high computational complexity. In [1], an algorithm called "Recursive Optimal Per-pixel Estimate" (ROPE) is proposed to compute the distortion by recursively calculating the first and second moments of each pixel due to compression, error concealment, and error propagation. This algorithm provides an accurate estimation of end-to-end distortion at the cost of a modest increase in computational complexity. Since the ROPE algorithm achieves substantial gains over competing methods, extensive work has been proposed based on the ROPE algorithm. For example, Eisenberg et. al estimate the variance of expected distortion by calculating the first four moments of each pixel and incorporates these moments to allocate channel resources [58]. In [4], the overall distortion is divided into several separable distortion items to reduce the computing complexity. Reibman extends the ROPE algorithm to a MD-split coder [59]. In [60], the authors estimate the expected end-to-end distortion to select multiple description modes on a frame basis.

All of these techniques only consider a simple network condition in which an average packet loss rate is assumed. However, [52] has shown that not only average packet loss rate but also the specific pattern of the loss affects the expected distortion; specifically, they prove that burst losses have a great impact on the distortion. Because of the likelihood of both random packet loss and burst losses in video communications over wireless networks, we propose a method which takes account this more complicated network condition for optimal mode selection to enhance the error resilience of delivered video [61]. Our method estimates the end-to-end distortion based on the ROPE algorithm considering both random and burst losses, and uses RD optimization for the

optimal mode selection. The method is applied in two cases. For single description video coding, we estimate the reconstructed pixel values due to random loss and burst losses, which results in a more precise estimation for the end-to-end distortion. When applying to RD-based mode selection, this method can achieve the optimal tradeoff between error resilience and coding efficiency under different random and burst loss rates, and outperforms ROPE algorithm over lossy networks. For multiple description video coding, we estimate the reconstructed pixel values by considering the network condition, error propagation and multiple state recovery, and select the mode that enhances the error robustness of MSVC.

4.2 Recursive Optimal per-Pixel Estimate and RD Based Mode Selection

Most of the video standards provide different intra and inter modes to encode a MB. For example, H.264 supports various coding modes such as Intra_16×16, Intra_4×4, Inter_SKIP, Inter_16×16, Inter_8×16, Inter_16×8, and Inter_8×8. In order to decide the best mode for each MB, a Lagrangian optimization technique is used to minimize the distortion subject to a rate constraint [62]. That is, the coding mode that minimizes the Lagrangian cost in the following equation is chosen to encode the MB,

$$\min_{mode}(J_{MB}) = \min_{mode}(D_{MB} + \lambda_{mode} R_{MB}) \tag{4.1}$$

where R_{MB} denotes the bits needed for coding the MB in the specific mode, which includes the bits for the MB header, the motion vector, the reference frame, and the transformed coefficients. D_{MB} represents the distortion of the MB, and λ_{mode} is the Lagrangian multiplier for the mode decision given by Eq. (4.2) in H.264,

$$\lambda_{mode} = 0.85 \times 2^{(QP-12)/3} \tag{4.2}$$

where QP is the quantization parameter of the MB.

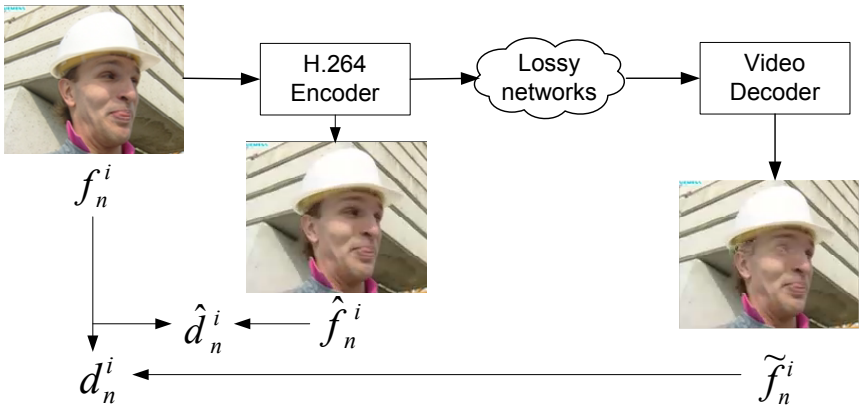


Figure 4.1: Video Coding and Transmission System

To determine the optimal mode for each MB, we need to estimate the distortion of the MB. In an H. 264 video encoder, the distortion is defined as the mean square error between the original video pixel value f_n^i and the encoded pixel value $\hat{f}_n^i$ as shown in Fig. 4.1. The coding mode chosen to encode the video is optimal for the compressed video without losses. However, for the video transmitted over lossy networks, the decoded pixel value $\tilde{f}_n^i$ suffers from packet losses and is not equal to $\hat{f}_n^i$. To select the optimal coding mode for the video that suffers losses, the encoder needs to estimate the distortion d_n^i between the

original pixel value f_n^i and the decoder-reconstructed pixel value $\tilde{f}_n^i$ as shown in Fig. 4.1.

Table 4.1 defines the notations used in the derivation of the distortion. The distortion of each MB is defined as the sum of the end-to-end distortion of the pixels in the MB,

$$D_{MB} = \sum_{i \in MB} d_n^i \tag{4.3}$$

Table 4.1: Notation

	Definitions
d_n^i	End-to-end distortion of pixel i in frame n
f_n^i	Original value of pixel i in frame n
$\hat{f}_n^i$	Encoder-reconstructed value of pixel i in frame n
$\tilde{f}_n^i$	Decoder-reconstructed value of pixel i in frame n (after error concealment)
$\hat{r}_n^i$	Quantized residue of pixel i in frame n (Inter mode)

The expected end-to-end distortion for the pixel f_n^i is given by

$$d_n^i = E[(f_n^i - \tilde{f}_n^i)^2] = (f_n^i)^2 - 2f_n^i E[\tilde{f}_n^i] + E[(\tilde{f}_n^i)^2] \tag{4.4}$$

Notice that the value of $\tilde{f}_n^i$ is a random variable at the encoder. In order to estimate the expected distortion d_n^i at the encoder, we need to calculate the first and second moments of $\tilde{f}_n^i$ for an intra and an inter MB separately.

4.3 Optimal Mode Selection for Single Description Coding with Random and Burst Losses

In [1], the authors develop the ROPE algorithm to recursively compute the first and second moments of $\tilde{f}_n^i$ based on the packet loss rate p and error

concealment method. We notice that the ROPE algorithm only considers a simple loss model, in which each packet may be lost with a packet loss rate p. While in wireless networks, the loss pattern is usually more complicated. The video packets may suffer from burst losses as well as random loss. [52] shows that the loss pattern has a significant impact on the distortion and burst losses generally cause a larger distortion than isolated losses. Therefore, we extend the ROPE algorithm with burst losses to better estimate the decoder-reconstructed pixel value for single description video coding.

When burst losses happen, the concealed pixel is further away from the last correctly received frame and it generally has a greater distortion. Therefore, we distinguish it from the concealed pixel value due to random loss. By separately estimating the concealed pixel value due to random loss and burst losses, we can more accurately calculate the end-to-end distortion at the encoder for optimal mode decision.

We assume that the temporal-copy error concealment is used to recover the lost video segment. That is, a lost MB is concealed by copying the previous correctly received MB in the corresponding position. The packet loss model in Section 3.4.1 is applied, in which three parameters need to be considered for the extended ROPE algorithm: burst loss rate p_b, burst length k (frames), and random packet loss rate p_r. Using the notations in Table 4.1, we calculate the first and second moments of $\tilde{f}_n^i$ in intra and inter modes respectively.

4.3.1 Pixel in an intra-coded MB

According to the packet loss model, each packet may experience three network conditions:

1. The packet is correctly received with probability $(1-p_b)(1-p_r)$. We thus have $\tilde{f}_n^i = \hat{f}_n^i$.

2. The packet suffers burst losses with probability p_b. This means that k consecutive frames are lost during the time interval Δt. The lost MB is then concealed by the co-located MB in the last correctly received frame. That is $\tilde{f}_n^i = \tilde{f}_{n-(n \bmod k)}^i$.

3. The packet encounters random loss with probability $(1-p_b)p_r$. Then the lost MB is recovered by copying the co-located MB in the previous frame. Therefore, we have $\tilde{f}_n^i = \tilde{f}_{n-1}^i$.

Based on the three cases, the first and second moments of $\tilde{f}_n^i$ in an intra-coded MB are calculated by,

$$\begin{aligned} E[\tilde{f}_n^i] = {} & (1-p_r)(1-p_b)\cdot(\hat{f}_n^i) + (1-p_b)p_r \cdot E[\tilde{f}_{n-1}^i] \\ & + p_b \cdot E[\tilde{f}_{n-(n \bmod k)}^i] \end{aligned} \tag{4.5}$$

$$\begin{aligned} E[(\tilde{f}_n^i)^2] = {} & (1-p_r)(1-p_b)\cdot(\hat{f}_n^i)^2 + (1-p_b)p_r \cdot E[(\tilde{f}_{n-1}^i)^2] \\ & + p_b \cdot E[(\tilde{f}_{n-(n \bmod k)}^i)^2] \end{aligned} \tag{4.6}$$

4.3.2 Pixel in an inter-coded MB

When the pixel is inter-coded, there are also three cases to estimate the decoder-reconstructed pixel value:

1. The packet is correctly received with probability $(1-p_b)(1-p_r)$. For an inter-coded pixel, we assume that pixel i is predicted from pixel j in the previous frame and the quantized residue is $\hat{r}_n^i$. Then the encoder reconstruction $\hat{f}_n^i$ is computed by adding the quantized residue to the prediction, that is, $\hat{f}_n^i = \hat{r}_n^i + \hat{f}_{n-1}^j$. Thus, the decoder-reconstructed pixel value is given by, $\tilde{f}_n^i = \hat{r}_n^i + \tilde{f}_{n-1}^j$.

2. The packet suffers burst losses with probability p_b. Similar to the intra-coded pixel, the pixel is concealed from the last correctly received frame and we have $\tilde{f}_n^i = \tilde{f}_{n-(n \bmod k)}^i$.

3. The packet encounters random loss with probability $(1-p_b)p_r$ and the pixel is concealed by the pixel in the previous frame: $\tilde{f}_n^i = \tilde{f_{n-1}^i}$.

Finally, the first and second moments of $\tilde{f}_n^i$ in an inter-coded MB are given by:

$$\begin{aligned} E[\tilde{f}_n^i] &= (1-p_r)(1-p_b)\cdot(\hat{r}_n^i + E[(\tilde{f}_{n-1}^j)]) \\ &\quad +(1-p_b)p_r \cdot E[\tilde{f}_{n-1}^i] + p_b \cdot E[\tilde{f}_{n-(n \bmod k)}^i] \qquad (4.7) \\ E[(\tilde{f}_n^i)^2] &= (1-p_r)(1-p_b)\cdot E[(\hat{r}_n^i + \tilde{f}_{n-1}^j)^2] \\ &\quad +(1-p_b)p_r \cdot E[(\tilde{f}_{n-1}^i)^2] + p_b \cdot E[(\tilde{f}_{n-(n \bmod k)}^i)^2] \qquad (4.8) \end{aligned}$$

Using Eqns. (4.5)-(4.8), we can recursively estimate the first and second moments of $\tilde{f}_n^i$ and calculate the overall end-to-end distortion for each MB. By

applying the RD-based mode selection method in Section 4.2, the optimal mode that provides a good trade-off between coding efficiency and error resilience for the specific random and burst loss rates is chosen.

4.4 Optimal Mode Selection for MSVC with Random and Burst Losses

From Section 2.2, we know that MSVC transmits two independently decodable descriptions over two different paths to reduce the loss of consecutive frames. Burst losses in one description only cause the loss of consecutive odd (even) frames, which can be well concealed by the even (odd) frames in the other description. On the other hand, burst losses can cause severe degradation to all the subsequent frames in SDC. Therefore, MSVC is more robust to burst losses than SDC. However, when MSVC experiences random packet loss, the distortion due to random loss not only propagates to subsequent frames in the same description, but may also affect frames in the other description because of multiple state recovery. In order to mitigate the error propagation due to random loss in MSVC, we propose a rate-distortion optimized mode selection method for MSVC, which adaptively encodes MBs in different modes to reduce the impact of error propagation.

The idea is similar to the ROPE method, except that MSVC uses multiple state recovery to conceal the error and the encoder needs to consider this during the estimation process. We assume that the refined error concealment methods presented in Chapter 3 are applied. We estimate the first and second moments

of $\tilde{f}_n^i$ by considering the packet loss rate p, and the multiple state recovery, and calculate the expected end-to-end distortion for each MB. When applying RD-based mode selection, the proposed method can better recover from random loss.

4.4.1 Pixel in an intra-coded MB

To compute the first and second moments of $\tilde{f}_n^i$ for an Intra MB, we need to consider the following scenarios:

1. The packet for f_n^i is correctly received with probability $1-p$ and thus we have $\tilde{f}_n^i = \hat{f}_n^i$.

2. The packet for f_n^i is lost and the neighbor group of blocks (GOB) is received with probability $p(1-p)$. In this case, we estimate the motion vector of lost pixel from one of the available neighbor MBs and use motion-compensated concealment to recover the lost pixel. We choose one frame as the reference from each description and get two reconstructed values $\tilde{f}_{n-1}^{j_1}$ and $\tilde{f}_{n-2}^{j_2}$. Then pixel $\tilde{f}_n^i$ is recovered from $\tilde{f}_{n-1}^{j_1}$ or $\tilde{f}_{n-2}^{j_2}$ depending on which reconstructed value is closer to $\hat{f}_n^i$, i.e. $\tilde{f}_n^i = \tilde{f}_{n-m}^{j_m}$, where $m = \arg\min_{x\in\{1,2\}}(\tilde{f}_{n-x}^{j_x} - \hat{f}_n^i)^2$.

3. The packet for f_n^i and the neighbor GOB are both lost with probability p^2. Then either $\tilde{f}_{n-1}^i$ or $\tilde{f}_{n-2}^i$ is used to conceal $\tilde{f}_n^i$. Thus, $\tilde{f}_n^i = \tilde{f}_{n-k}^i$, where $k = \arg\min_{x\in\{1,2\}}(\tilde{f}_{n-x}^i - \hat{f}_n^i)^2$.

Based on the above cases, we can calculate the first and second moments of $\tilde{f}_n^i$ in an intra MB by Eqs. (4.9) and (4.10),

$$E[\tilde{f}_n^i] = (1-p)(\hat{f}_n^i) + p(1-p)E[\tilde{f}_{n-m}^{j_m}] + p^2 E[\tilde{f}_{n-k}^i] \tag{4.9}$$

$$E[(\tilde{f}_n^i)^2] = (1-p)(\hat{f}_n^i)^2 + p(1-p)E[(\tilde{f}_{n-m}^{j_m})^2] + p^2 E[(\tilde{f}_{n-k}^i)^2] \tag{4.10}$$

$$\text{where} \quad m = \arg\min_{x\in\{1,2\}}(E[\tilde{f}_{n-x}^{j_x}] - \hat{f}_n^i)^2, \quad \text{and} \quad k = \arg\min_{x\in\{1,2\}}(E[\tilde{f}_{n-x}^i] - \hat{f}_n^i)^2$$

4.4.2 Pixel in an inter-coded MB

For MSVC, the odd frame is predicted from the previous odd frames and the even frame is predicted from the previous even frames. Therefore, the quantized residue $\hat{r}_n^i = \hat{f}_n^i - \hat{f}_{n-2}^j$ for MSVC, where pixel i in frame n is predicted from pixel j in frame $n-2$. We assume that $j_m (m = 1, 2)$ is the pixel corresponding to the estimated concealment motion vector for pixel i in frame $n-m$. Then we can calculate the first and second moments of $\tilde{f}_n^i$ according to the three cases similar to those in Section 4.4.1,

$$E[\tilde{f}_n^i] = (1-p)(\hat{r}_n^i + E[\tilde{f}_{n-2}^j]) + p(1-p)E[\tilde{f}_{n-m}^{j_m}] + p^2 E[\tilde{f}_{n-k}^i] \tag{4.11}$$

$$E[(\tilde{f}_n^i)^2] = (1-p)E[(\hat{r}_n^i + \tilde{f}_{n-2}^j)^2] + p(1-p)E[(\tilde{f}_{n-m}^{j_m})^2] + p^2 E[(\tilde{f}_{n-k}^i)^2] \tag{4.12}$$

$$\text{where} \quad m = \arg\min_{x\in\{1,2\}}(E[\tilde{f}_{n-x}^{j_x}] - \hat{f}_n^i)^2, \quad \text{and} \quad k = \arg\min_{x\in\{1,2\}}(E[\tilde{f}_{n-x}^i] - \hat{f}_n^i)^2$$

4.5 Performance Evaluation of RD-optimized Mode Selection with Random and Burst Losses

In this section, we evaluate the performance of the proposed methods introduced in Section 4.3 and Section 4.4 under different network conditions.

We implement our proposed methods by modifying H.264 reference software JM13.2. We use the temporal copy method in the implementation. That is, the lost MB is concealed by copying the co-located MB in the last correctly received frame. The following four approaches are implemented for comparison:

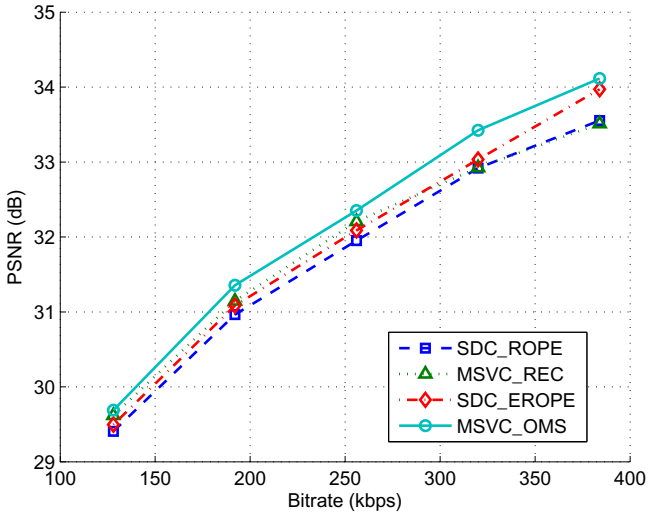
- **SDC_ROPE**: The video sequence is coded into a single description with ROPE proposed in [1] and transmitted over one path over the network.
- **MSVC_REC**: The video sequence is coded into two descriptions using MSVC and transmitted over two independent paths over the network. At the decoder, refined error concealment introduced in Chapter 3 is applied to decode the corrupted video.
- **SDC_EROPE**: The video sequence is coded into a single description with extended ROPE method proposed in Section 4.3 and transmitted through one path in the network.
- **MSVC_OMS**: The video sequence is coded into two descriptions using MSVC with the optimal mode selection proposed in Section 4.4. Then

the encoded bitstream is transmitted over two paths and decoded using the refined error concealment method for MSVC proposed in Chapter 3.

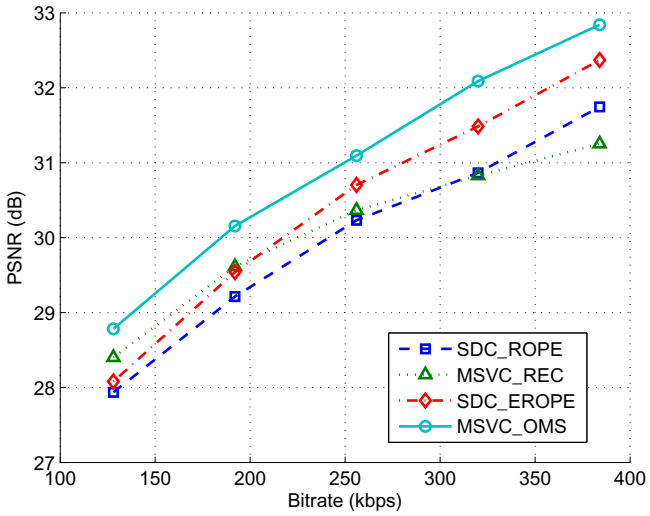
We use the packet loss model introduced in Section 3.4.1 to simulate random and burst losses in the networks. We evaluate six video sequences including Carphone, Claire, Foreman, Hall-monitor, Mother-daughter, and News. Each sequence consists of 300 frames at QCIF format. The sequences are encoded as IPPP format with $GOP = 30$ at 30 fps and each frame is packetized to 4 RTP packets.

First, we show the average PSNRs of SDC_ROPE, MSVC_REC, SDC_EROPE and MSVC_OMS under different bitrates for two sets of network conditions in Fig. 4.2. We see that MSVC_OMS achieves about 0.43 dB, 0.31 dB, and 0.25 dB gains under different bitrates at a 4% packet loss rate compared to SDC_ROPE, MSVC_REC, and SDC_EROPE, respectively. At a 8% packet loss rate, the gains of MSVC_OMS are 1.00 dB, 0.90 dB, and 0.56 dB compared to SDC_ROPE, MSVC_REC, and SDC_EROPE. We see that MSVC_OMS achieves the best average PSNR performance compared to other three methods under different bitrates at different packet loss rates.

We have shown the objective performance of four methods in Fig. 4.2. Now we investigate the video quality of the four methods for multiple users. Figure 4.3 shows $\text{PSNR}_{r,f}$ of SDC_ROPE, and MSVC_REC, SDC_EROPE, and MSVC_OMS for Foreman sequence under network condition ($p_r = 4\%, p_b = 4\%, k = 5$). Here, p_r is the random loss rate, p_b is the burst loss rate, and k is the burst length (in frames). The average PSNR of these four methods are 30.23 dB, 30.37 dB, 30.70 dB, and 31.09 dB respectively. Figure 4.3(a) presents the



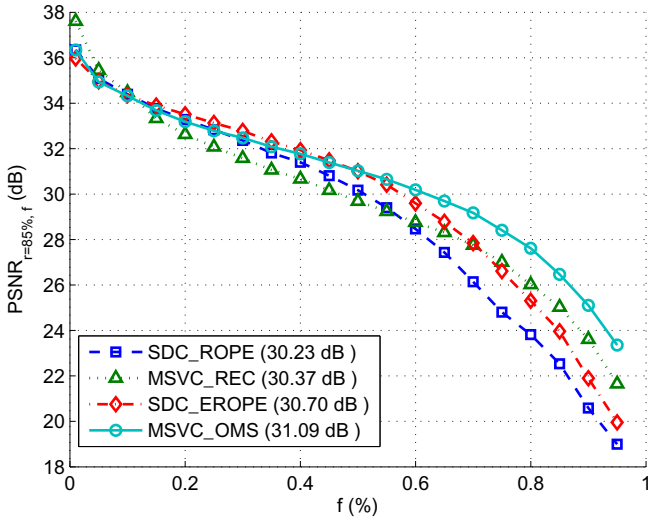
(a) $p_r = 2\%, p_b = 2\%, k = 5$



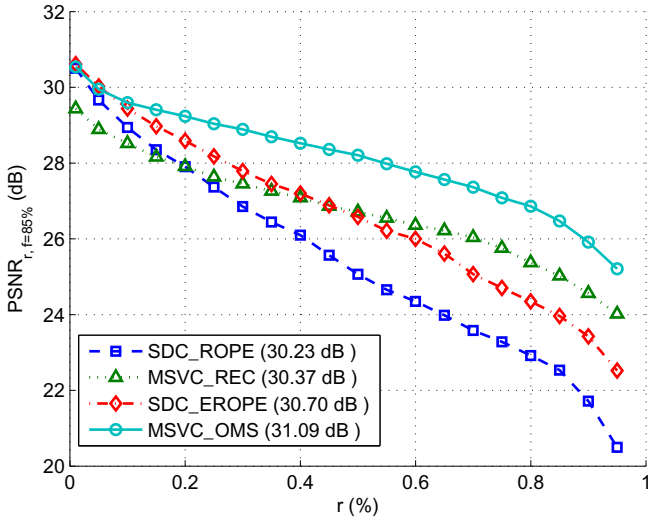
(b) $p_r = 4\%, p_b = 4\%, k = 5$

Figure 4.2: Average PSNR vs Bitrate for SDC_ROPE, MSVC_REC, SDC_EROPE and MSVC_OMS, Foreman Sequence at 30 fps

$\text{PSNR}_{r,f}$ values with a fixed $r = 85\%$. Compared to SDC_ROPE, MSVC_REC, and SDC_EROPE, MSVC_OMS has the fewest number of low-quality frames in 85% of the realizations. For example, as shown in Fig. 4.3(a), about 25% of



(a) $\text{PSNR}_{r=85\%,f}$



(b) $\text{PSNR}_{r,f=85\%}$

Figure 4.3: $\text{PSNR}_{r,f}$ for Foreman sequence at 30 fps, 256 kbps, $p_r = 4\%$, $p_b = 4\%$, $k = 5$

frames in 85% of the realizations for SDC_ROPE have a PSNR lower than 25 dB, while fewer than 10% of frames in 85% of the realizations for MSVC_OMS achieve a PSNR lower than 25 dB. We also see that SDC_EROPE has consis-

tently higher PSNR_f values than SDC_ROPE for a fixed r, which means that for 85% of users, SDC_EROPE constantly provides better video quality than SDC_ROPE. Furthermore, the $\text{PSNR}_{r=85\%,f}$ curve of MSVC_REC crosses over the $\text{PSNR}_{r=85\%,f}$ curves of SDC_ROPE and SDC_EROPE at $f = 55\%$ and $f = 70\%$ respectively. It means that the PSNR variance of MSVC_REC is smaller than SDC_ROPE and SDC_EROPE, which provides better perceptual video quality for human viewers.

Figure 4.3(b) plots $\text{PSNR}_{r,f}$ of SDC_ROPE, MSVC_REC, SDC_EROPE, and MSVC_OMS with fixed $f = 85\%$. In the figure, we see that MSVC_OMS has the highest $\text{PSNR}_{r,f=85\%}$ under most values of r. This means that MSVC_OMS can guarantee a higher PSNR than SDC_ROPE, SDC_EROPE, and MSVC_REC for 85% of frames in all of the realizations. For example, the PSNRs guaranteed for 85% of the frames in 85% of the realizations for SDC_ROPE, MSVC_REC, SDC_EROPE, and MSVC_OMS are 22.53 dB, 25.02 dB, 23.96 dB, and 26.47 dB respectively. This indicates that MSVC_OMS guarantees a higher video quality for a user in multiple channel uses (r%) or provides better video experience for multiple users in the network. Based on Fig. 4.3, we see that SDC_EROPE has better performance than SDC_ROPE, and MSVC_OMS achieves the best perceptual quality for multiple users.

Table 4.2 and Table 4.3 present the average PSNR and $\text{PSNR}_{r=85\%,f=85\%}$ results for SDC_ROPE, MSVC_REC, SDC_EROPE and MSVC_OMS for different video sequences. The results in Table 4.2 show that MSVC_OMS consistently provides PSNR gains for various sequences. One exception is for News sequence at packet loss rate $p_b = 2\%, p_r = 2\%, k = 5$. MSVC_OMS has lower

Table 4.2: Average PSNR for Different Video Sequences, 30 fps, 256 kbps

Sequence	$p_b = 2\%, p_r = 2\%, k = 5$			
	SDC_ROPE	MSVC_REC	SDC_EROPE	MSVC_OMS
Carphone	33.23	33.21	33.27	33.88
Claire	41.37	42.31	41.20	42.66
Foreman	31.95	32.21	32.09	32.35
Hall-monitor	38.40	38.46	38.40	39.09
Mother-daughter	38.90	39.20	38.98	39.45
News	37.84	37.08	37.52	37.31
	$p_b = 4\%, p_r = 4\%, k = 5$			
Carphone	31.25	31.02	31.33	32.38
Claire	38.68	40.42	38.11	41.17
Foreman	30.23	30.37	30.70	31.09
Hall-monitor	35.66	36.52	35.58	37.72
Mother-daughter	37.25	37.89	37.59	38.38
News	35.26	35.20	35.28	35.80

Table 4.3: $\text{PSNR}_{r=85\%,f=85\%}$ for Different Video Sequences, 30 fps, 256 kbps

Sequence	$p_b = 2\%, p_r = 2\%, k = 5$			
	SDC_ROPE	MSVC_REC	SDC_EROPE	MSVC_OMS
Carphone	23.97	27.86	24.51	28.50
Claire	27.22	37.31	28.85	38.84
Foreman	24.29	27.99	25.92	28.64
Hall-monitor	25.76	32.58	25.56	33.77
Mother-daughter	30.69	35.91	31.75	36.82
News	25.21	30.58	25.56	31.39
	$p_b = 4\%, p_r = 4\%, k = 5$			
Carphone	21.63	25.17	21.83	26.31
Claire	24.40	32.88	25.74	35.28
Foreman	22.53	25.02	23.96	26.47
Hall-monitor	22.17	29.21	22.92	29.60
Mother-daughter	26.82	33.18	29.00	34.40
News	22.67	26.66	22.90	26.98

average PSNR, probably because the gain that comes from error resilience of MSVC_OMS is not enough to compensate the coding efficiency loss due to reduced correlation between adjacent frames in the same description. However, for the same case, the $\text{PSNR}_{r=85\%,f=85\%}$ value of MSVC_OMS is 6.2 dB higher than SDC_ROPE, which means even though MSVC_OMS has a lower average PSNR value than SDC_ROPE, it has fewer number of bad-quality frames and provides better perceptual quality for multiple users.

4.6 Discussions

4.6.1 Complexity Considerations

In this section, we compare the cost of computational complexity and storage of our proposed approaches with ROPE. We know that ROPE leads to a modest increase in computational complexity, which is mostly introduced by calculating the two moments of $\tilde{f}_n^i$ for both intra mode and inter mode for each pixel. When comparing EROPE to ROPE, we know that the concealed pixel value caused by burst losses is estimated separately and it introduces 4 more addition/multiplication operations for each pixel in an intra-coded/inter-coded MB. Since the error concealment is the same regardless of the coding mode of the MB, the total number of extra addition/multiplication operations is 6 for each pixel. For MSVC_OMS, the extra operations come from the selection of multiple state recovery and it is the same for both intra-coded MB and inter-coded MB. Therefore, MSVC_OMS requires 8 more extra addition/multiplication operations for each pixel than ROPE. Based on the above analysis, we see that the

computational complexity of EROPE and MSVC_OMS is in the same order of ROPE. Furthermore, all the additional complexity occurs only at the encoder.

For storage cost, we see that ROPE only needs to store the two moments of each pixel in the previous frame, while EROPE stores the two moments of previous k frames and MSVC_OMS stores the two moments of the previous two frames. This extra storage cost introduced by EROPE and MSVC_OMS is negligible in most applications.

4.6.2 Mismatch of Network Conditions

Our approaches assume that the network conditions are known at the encoder and are used as the coding parameters. We need to analyze the situation that mismatch happens between the assumed network condition and actual condition in the network. There are two cases to be considered: either the assumed packet loss rate is lower or higher than the actual loss rate in the network. For the first case, the distortion caused by packet loss for SDC_ROPE, SDC_EROPE, and MSVC_OMS is all underestimated; nevertheless, SDC_EROPE and MSVC_OMS are still more robust to packet loss than SDC_ROPE. For the second case, the distortion is overestimated and it may introduce unnecessary redundancy for error resilience. The worst mismatch in this case is that the network is error-free and the decoder reconstruction is equal to the encoder reconstruction. For example, when the assumed network condition is $p_r = 4\%, p_b = 4\%, k = 5$, the average PSNRs at the encoder for SDC_ROPE, MSVC_REC, SDC_EROPE, and MSVC_OMS are 34.35 dB, 34.54 dB, 33.72 dB, and 33.72 dB, respectively. When no packet loss happens in

the network, the average PSNRs at the decoder are the same as above PSNRs, which all represent quite good video quality. As we discussed in Chapter 2, for PSNRs higher than a certain threshold, increasing PSNR does not help to enhance the perceptual quality [20], thus the loss in coding efficiency does not affect the perceptual video quality much. Our proposed approaches start to achieve gains when the video transmission suffers packet losses over the network.

Chapter 5

Routing-aware Reference Frame Selection for MSVC

Supporting video transmission over error-prone mobile ad-hoc networks is becoming increasingly important as these networks become more widely deployed. In this chapter, we propose a routing-aware multiple description video coding approach to support video transmission over mobile ad-hoc networks with multiple path transport. This approach uses ad-hoc routing messages available in the standard routing protocols to estimate the packet loss and then select the reference frames accordingly. We first explore the relationship between packet losses and routing messages. We build a model to estimate the packet loss probability of each packet according to the routing messages received by the transmitter and the transmission delay determined by the MAC layer access mechanism and the network parameters. Based on this model, we then estimate the frame loss probability, and apply a threshold-based algorithm to select the reference frames to mitigate error propagation. Unlike common

reference picture selection (RPS) work [63, 64], our approach does not require any extra channel feedback but retrieves information from normal routing messages. Experiments are conducted using the QualNet simulator that accounts for node mobility, channel properties, MAC operation, multipath routing, and traffic type. We evaluate the estimation accuracy of the proposed estimation model and investigate the objective and perceptual video quality for multiple users.

This chapter is organized as follows. We review the existing work on multiple description coding with path diversity in Section 5.1. In Section 5.2, we describe the architecture of our proposed routing-aware MDC system. Section 5.3 presents the proposed packet loss estimation method based on routing messages and network conditions and Section 5.4 discusses our reference frame selection algorithm for MDC using the estimated packet loss. The simulation setup for the routing-aware MDC system is introduced in Section 5.5. In Section 5.6, we analyze the performance of our proposed method under different network conditions.

5.1 MDC with Path Diversity for Video over Mobile Ad-hoc Networks

Multiple description coding (MDC) has been shown to be a promising technique for video transmission over lossy networks[5]. With MDC, a video sequence is encoded into multiple descriptions such that each description can be used to reconstruct the video with low but acceptable video quality. When more

number of descriptions are received for reconstruction, higher video quality can be achieved. As long as all descriptions are not lost simultaneously, acceptable quality can be maintained. In order to reduce the likelihood of simultaneous loss of descriptions, different descriptions are transmitted through different paths. This is referred as MDC with multiple path transport (MPT). MPT helps to reduce the possibility of simultaneous loss of different descriptions and enables load balancing in networks. Many studies show that combining MDC with MPT leads to substantial performance gains for video transmission over these networks [13, 60, 65, 66].

The research in this area can be generally divided into two categories. One category studies the effectiveness of MDC methods based on a specific network model with path diversity [60, 65, 66]. In [65], the authors proposed a MDC method based on the lapped orthogonal transform and examined the performance on a two-path system with the same capacity and error characteristics. An adaptive MD mode selection approach is proposed in [60] to adapt to the network conditions as well as to the video characteristics. This approach selects the optimal MD mode by calculating the end-to-end distortion based on the Gilbert packet loss model. In [66], Mao et. al. compared feedback based reference picture selection, layered coding, and MDC schemes with multipath transport and found that MDC is preferable when a feedback channel cannot be set up.

The other category addresses the path selection and rate allocation problem for MDC given a particular MDC scheme [67–71]. Begen et. al. proposed a multi-path selection method that chooses a set of paths maximizing the overall

quality at the client based on the network parameters, media characteristics and application requirements [67]. The authors in [68] formulated a routing optimization problem that minimizes the application layer video distortion and provided a genetic-algorithm based approach to compute two disjoint paths for video transmission. In [69], the authors formulated the video distortion as a function of network layer behavior and proposed a branch-and-bound framework to produce optimal solutions. Different metrics used for the path selection for MDC are discussed in [70], and a practical interference aware distributed routing protocol is proposed. Zhou et al. proposed a joint routing and rate control algorithm to distribute video with optimal end-to-end quality of all users [71].

Our proposed method falls into the first category; however, instead of assuming that two node-disjoint paths with the same error characteristics are available or the set of paths is given, we consider multipath routing in a more practical network and utilize the route messages to select the proper reference frames. Our work is inspired by the reference picture selection (RPS) methods proposed in [63, 64]. Most of the RPS work assumes an extra feedback control channel from the video receiver to the sender, and the receiver thus sends an ACK/NACK for every video packet [63]. Such an approach can lead to extra overhead and cost, especially in a large network. Our routing-aware MDC method, on the other hand, does not require any additional control packets or an extra channel connection in the network. We only extract and utilize the information embedded in typical routing messages, thus saving network bandwidth.

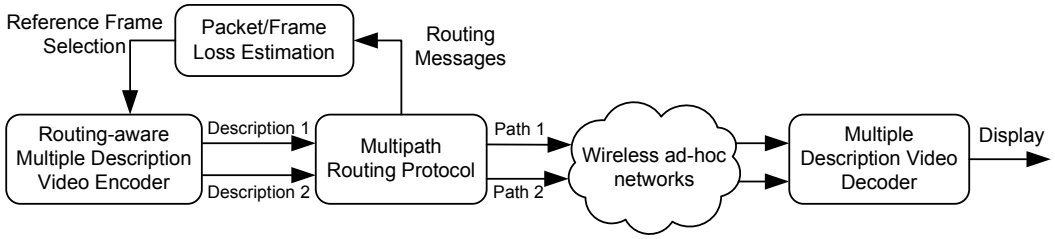


Figure 5.1: System Architecture of The Proposed System Using Routing-aware Multiple Description Coding and Multipath Routing

5.2 Routing-aware Multiple Description Video Coders

The system architecture of the proposed system is shown in Fig. 5.1. The routing-aware multiple description video encoder generates two video descriptions based on the MSVC method [13] and routing-aware reference frame selection. The two descriptions are transmitted through two paths established by the multipath routing protocol. At the receiver, the descriptions are decoded by the MD video decoder, in which the refined error concealment method proposed in [72] is applied to reconstruct the video sequence from received descriptions. We discuss the MD video encoder/decoder and the multipath routing protocol in this section and present our packet loss estimation approach in Section 5.3.

5.2.1 Multiple Description Video Encoder/Decoder

Multiple description video coding (MDC) is an effective approach to enhance the error resilience of video transmission over lossy networks. The general idea is to encode the video sequence into several descriptions with equal impor-

tance. Each description can be decoded independently or combined with other descriptions for reconstruction. In general, the reconstructed video achieves better video quality when more descriptions are received.

Among the many proposed MDC algorithms [5], multiple state video coding (MSVC) proposed by Apostolopoulos in [13] is a very popular method since it is easy to implement and compatible with different video standards. Thus, we apply MSVC to our MD video encoder. At the encoder, the video sequence is temporally downsampled into two sub-sequences consisting of odd and even frames, and the odd and even frames are encoded as two descriptions using an H.264 encoder. During the encoding process, we use routing messages from the routing protocol to help the encoder select the reference frames. The details are presented in Section 5.4.

At the decoder, we utilize the MSVC decoder with the refined error concealment method as proposed in [72]. When the decoder receives the corrupted descriptions, it decodes the correctly received MBs and conceals the lost MBs with the refined MB concealment method that considers the information from both descriptions for better recovery. The refined intra MB concealment reconstructs the lost MBs in the intra frames by using the temporal correlation between adjacent intra frames in two descriptions, while the refined inter MB concealment uses an additional reference list to perform the motion-compensated concealment. Finally, the concealed descriptions are interleaved to achieve the final reconstruction.

5.2.2 Multipath Routing Protocol

As mentioned in Section 5.1, combining MDC with multipath transport is an appealing approach because it provides error resilience as well as load balancing for video transmission over networks. To support MDC with path diversity, a multipath routing protocol is required to build multiple paths between the source and destination nodes through the ad-hoc network.

Many multipath routing protocols have been proposed to support multipath transport in wireless ad-hoc networks [73–76]. In [73], the authors proposed a multipath extension of dynamic source routing (DSR) [77], in which a set of alternate link-disjoint routes are maintained. Another extension of DSR called split multipath routing (SMR) is proposed in [74]. It focuses on building and maintaining multiple maximally disjoint paths. AOMDV [75] and AODVM [76] are two multipath protocols extended from the ad-hoc on-demand distance vector (AODV) routing protocol, in which AOMDV computes multiple loop-free and link-disjoint paths [75] and AODVM finds multiple node-disjoint paths [76].

Although these on-demand multipath routing protocols have different optimization criteria to establish routes, they all consist of two basic mechanisms: route discovery and route maintenance. A route discovery process is triggered when a source node needs a route to transmit packets to a destination node. A route request (RREQ) message is flooded to the entire network to find the routes. When the RREQ reaches the destination node, a route reply (RREP) message is sent back to the source node to build a new route. Route maintenance deals with the situation that a route becomes worse or even broken.

When a route breaks, the node that detects the link failure sends a route error (RERR) message to the source node. Once the source node receives the RERR, either a new route is built from the route table or a route discovery is initiated to reconstruct a new route.

We notice that the routing discovery and maintenance mechanisms in most of the routing protocols provides feedback concerning the network conditions. This inspires us to utilize these feedback messages to estimate the packet losses in the network and to adapt the video coding accordingly. In this paper, we implement SMR as our multipath routing protocol due to its popularity and simplicity [78]. However, our solution is not limited to this particular protocol and can be extended to other multipath routing protocols.

5.3 Packet Loss Estimation via Routing Messages

In this section, we present how to use routing messages to estimate the packet losses in the network. Based on the routing mechanisms, a RERR message is initiated when the MAC layer fails all retransmission attempts to transmit a packet to the next hop destination. This RERR indicates that a link becomes unreliable and the packets transmitted through this link suffer a high packet loss rate. Before the source node receives the RERR, video packets sent from the source node are still transmitted through this error-prone link and are susceptible to losses. When the source receives the RERR, it either reconstructs the route from the route cache or initiates the route recovery process to find a

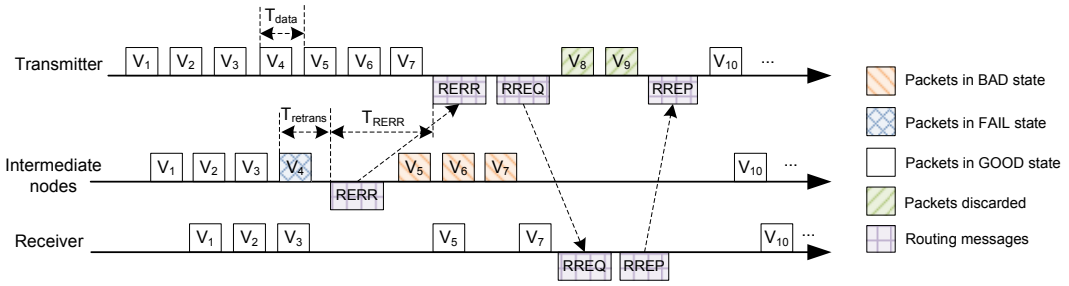


Figure 5.2: An Example to Illustrate The Packet Losses in The Network and The Corresponding Routing Messages

new route. Packets scheduled to be transmitted in the broken route during the route recovery process are discarded and marked as lost.

In our previous work [79], we use a simple method to utilize the routing messages, that is, every time a route error (RERR) message is received by the source node, we assume that the previously transmitted packet is lost. However, due to the transmission delay of the video packets and routing messages, a RERR may indicate possible losses of several previously transmitted video packets. Therefore, here we propose a model to estimate the packet loss probability of the packets sent through an unreliable link.

Figure 5.2 illustrates how the RERR message correlates to the packet losses in the network. As shown in Fig. 5.2, a RERR is initiated at the intermediate node when video packet v_4 exhausts all retransmission attempts and still fails to transmit to the next hop destination. We define the retransmission delay of this packet as T_{retrans}. After time T_{RERR}, the source node receives the RERR and stops transmitting video packets through the unreliable link. We see that packets v_5, v_6, v_7 sent during time period $T_{\text{retrans}} + T_{\text{RERR}}$ are still transmitted through the unreliable link and are very susceptible to packet loss.

We assume that anytime the source receives a RERR, the preceding video packets sent from the source follow the same packet loss distribution under the same network conditions. Therefore, we denote $\Pr(n)$ as the packet loss probability of the n^{th} preceding packet sent from the source before the source node receives a RERR. Our main goal is to model $\Pr(n)$ and utilize it to determine the potential corrupted frames. Due to the random delay between link failure and RERR reception at the source, the n^{th} preceding packet before RERR can be sent at a time before, right at, or after the link failure happens. We use three states to represent these three cases: GOOD means the packet is sent before the link failure, FAIL means the packet fails to transmit and triggers RERR, and BAD means the packet is sent after the link failure. According to our above analysis, we define $\Pr(n)$ as

$$\Pr(n) = \lambda_g \cdot p_g(n) + \lambda_f \cdot p_f(n) + \lambda_b \cdot p_b(n) \tag{5.1}$$

where λ_g, λ_f, and λ_b represents the packet loss probability in GOOD, FAIL, or BAD state respectively, and $p_g(n)$, $p_f(n)$, and $p_b(n)$ denotes the probability of the n^{th} preceding packet in these three states, respectively. In the following, we estimate the state probability distribution and packet loss probabilities in these three states.

5.3.1 Estimation of State Probability Distribution

The state of a video packet depends on the delay of the link failure feedback and the transmission interval of video packets. For example, as shown in Fig. 5.2, v_4 is the packet that triggers RERR and hence is in FAIL state. The packets

sent before v_4 (e.g. v_3) are in GOOD state while the packets sent after v_4 are in BAD state. Therefore, we can compare the video packet transmission interval T_{data} and the delay of the link failure feedback T_{delay} to determine the state of the packets sent before receiving the RERR by

$$\begin{cases} p_g(n) = p\left(T_{\text{delay}} \leq (n-1)T_{\text{data}}\right) \\ p_f(n) = p\left(nT_{\text{data}} \geq T_{\text{delay}} > (n-1)T_{\text{data}}\right) \\ p_b(n) = p\left(T_{\text{delay}} > nT_{\text{data}}\right) \end{cases} \tag{5.2}$$

We can calculate the video packet interval T_{data} by

$$T_{\text{data}} = L/R_t \tag{5.3}$$

where R_t is the transmission bitrate of the video sequence and L is the payload size. Then in order to calculate Eq. (5.2), we need to estimate the probability distribution of T_{delay}.

As shown in Fig. 5.2, we see that T_{delay} consists of two parts: the retransmission delay of a packet that fails all retransmission attempts (denoted as T_{retrans}) and the time period to transmit the RERR to the source (denoted as T_{RERR}). So we have

$$T_{\text{delay}} = T_{\text{retrans}} + T_{\text{RERR}} \tag{5.4}$$

The values of both T_{retrans} and T_{RERR} depend on the MAC layer access mechanism. The basic access method of the IEEE 802.11 MAC layer is the distributed coordination function (DCF) based on the carrier sense multiple access with collision avoidance (CSMA/CA) scheme [80]. The DCF method provides a basic access mechanism and an optional RTS/CTS access mechanism. In this paper, our estimation is based on the basic access mechanism. Our

method can also be applied to the RTS/CTS mechanism. Next, we estimate T_{retrans} and T_{RERR} based on the basic 802.11 DCF mechanism.

5.3.1.1 Estimation of T_{RERR}

We first estimate the time period to transmit the RERR to the source, T_{RERR}, by

$$T_{\text{RERR}} = n_{\text{hop}} \cdot T_C \tag{5.5}$$

where n_{hop} is the average number of hops to transmit RERR to the source, and T_C is the transmission time for a successful RERR transmission. For the basic 802.11 access mechanism, we have

$$T_C = T_{\text{DIFS}} + T_H + T_{\text{ctl}} + T_{\text{SIFS}} + T_{\text{ACK}} \tag{5.6}$$

where T_{DIFS} is the DIFS time, T_H represents the transmission time of MAC and PHY header, T_{ctl} is the transmission time of RERR payload, T_{SIFS} is the SIFS time, and T_{ACK} denotes the ACK transmission time.

5.3.1.2 Estimation of T_{retrans}

We then estimate the transmission delay T_{retrans} of a packet that fails to transmit from the current station to the next hop destination after exhausting all retransmission attempts. As shown in Fig. 5.3, each transmission period consists of a defer access and a backoff process. The transmission procedure starts when the station senses an idle distributed inter-frame space (DIFS) and invokes a backoff procedure. The backoff time is uniformly chosen in the range of $[0, CW]$, where CW is the current contention window (CW) size. Then the

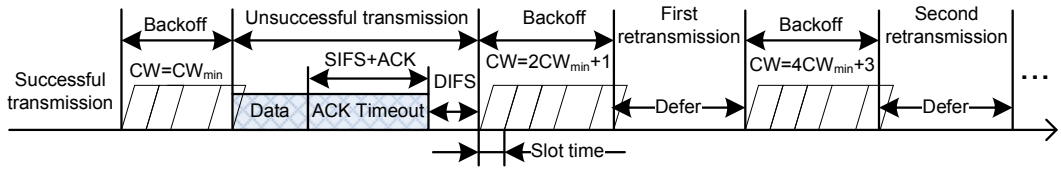


Figure 5.3: Packet Retransmission Procedure Based on Basic Access Mechanism

station sends out the video packet. If the transmitting station does not receive the acknowledgment (ACK) within the ACK timeout interval, the station concludes that the transmission has failed and invokes a retransmission process until the retransmission limit is reached. Note that CW takes an initial value of $CW_{\min}$ and exponentially increases after each unsuccessful transmission, until it reaches the maximum CW size of $CW_{\max}$.

Based on the above analysis, the transmission delay of a packet that fails all retransmission attempts is

$$T_{\text{retrans}} = mT_D + T_{\text{backoff}} \tag{5.7}$$

where m is the retransmission limit, T_D is the time period of a defer access, and T_{backoff} is the overall backoff time.

Similar to Eq. (5.6), we have T_D

$$T_D = T_{\text{DIFS}} + T_H + T_{\text{data}} + T_{\text{SIFS}} + T_{\text{ACK}} \tag{5.8}$$

where T_{data} is the transmission time of the data payload.

The overall backoff time is a random variable that is the sum of a series of independent random variables uniformly distributed in the range of $[0, W_i] \cdot T_{\text{slot}}$

and W_i is the CW size in the i^{th} retransmission defined by

$$W_i = \begin{cases} 2^i \cdot (CW_{\min}+1) - 1 & i \leq m' \\ 2^{m'} \cdot (CW_{\min}+1) - 1 & i > m' \end{cases} \tag{5.9}$$

where $m' = \log_2 \frac{(CW_{\max}+1)}{(CW_{\min}+1)}$.

We define T_{B_i} as the backoff time in the i^{th} retransmission, so we have $T_{B_i} \sim U(0, W_i \cdot T_{\text{slot}})$, where $U(0, W_i \cdot T_{\text{slot}})$ represents an uniform distribution in the range $[0, W_i] \cdot T_{\text{slot}}$. Thus the overall backoff time is

$$T_{\text{backoff}} = \sum_{i=0}^{m-1} T_{B_i} \sim U_{\text{s}}(0, \sum_{i=0}^{m-1} W_i \cdot T_{\text{slot}}) \tag{5.10}$$

where $U_{\text{s}}(\cdot)$ represents the probability distribution of the overall backoff time T_{backoff}, which is the sum of m uniform random variables. We use $P_s(t)$ to represent the CDF of T_{backoff}, *i.e.* the probability that T_{backoff} is shorter than time t is represented by $P_s(t)$.

Finally, based on Eqs. (5.2)-(5.10), we have the state probability distribution by

$$\begin{cases} p_g(n) = P_s(\Delta T) \\ p_f(n) = P_s(T_{\text{data}} + \Delta T) - P_s(\Delta T) \\ p_b(n) = 1 - P_s(T_{\text{data}} + \Delta T) \end{cases} \tag{5.11}$$

where $\Delta T = (n-1)T_{\text{data}} - T_{\text{RERR}} - mT_D$.

5.3.2 Estimation of Packet Loss Probability λ_g, λ_f, and λ_b

λ_g refers to the packet loss rate of a good link, in which the ACK is received to indicate a successful transmission. Therefore, we assume $\lambda_g = 0$. λ_b is

defined as the packet loss rate of an unreliable link, which is the probability that the video packet does not reach the next hop destination successfully. λ_f is the packet loss rate for the video packet that fails all transmission attempts and triggers the RERR. Based on the MAC layer mechanism, we know that each time a video packet fails a transmission, it means either the video packet fails to transmit to the next hop destination or the ACK message is not received by the transmitter. Thus, λ_f is the conditional packet loss probability for the video packet that fails all transmission attempts.

Let A_0 denote the event that the video packet is lost and A_1 denote the event that the video packet fails all transmission attempts. We assume that each transmission is independent and the loss probability of a video packet and an ACK for an unreliable link are p_{data} and p_{ACK} respectively. Then we have $p(A_0) = p_{\text{data}}^m$ and $p(A_1) = [p_{\text{data}} + (1 - p_{\text{data}}) \cdot p_{\text{ACK}}]^m$, where m is the retransmission limit. Finally, λ_f and λ_b are represented by

$$\begin{aligned} \lambda_f &= p(A_0|A_1) = \frac{p(A_1|A_0) \cdot p(A_0)}{p(A_1)} \\ &= \frac{p_{\text{data}}^m}{[p_{\text{data}} + (1 - p_{\text{data}}) \cdot p_{\text{ACK}}]^m} \qquad (5.12) \\ \lambda_{\text{b}} &= p(A_0) = p_{\text{data}}^m \qquad (5.13) \end{aligned}$$

By Eqs. (5.12) and (5.13), we have

$$\frac{\lambda_b}{\lambda_f} = [p_{\text{data}} + (1 - p_{\text{data}}) \cdot p_{\text{ACK}}]^m \leq 1 \qquad (5.14)$$

i.e. λ_f is generally larger than λ_b.

5.4 Routing-aware Reference Selection for MSVC Based on Packet Loss Estimation

Given the packet loss probability estimated from routing messages, we seek to design a routing-aware MDC method that can improve the error resilience of the reconstructed video. Since error propagation is probably the most important problem for video transmitted over error-prone channels [52], our design goal is to reduce error propagation caused by the video packet losses. Our design achieves this goal by using the reference frame selection technique to reduce error propagation, *i.e.*, select proper reference frames that do not suffer packet losses. For every video frame that may be corrupted, our design can estimate the frame corruption probability and avoid using this frame as a reference frame if the corruption probability is higher than a certain threshold. In the following, we describe the details of the frame corruption estimation and the reference selection algorithm for our MDC method.

In Section 5.3, we discussed the process to estimate the packet loss probability of each transmitted packet based on the routing messages. Based on the received RERR message, we estimate the packet loss probability of the n^{th} preceding packet sent from the source as $\Pr(n)$, which corresponds to the packet loss probability of the video packet with index v_i. We now use the estimated packet loss probabilities to determine the frame corruption probability of each frame by

$$p(f_k) = 1 - \prod_{v_i \in f_k} (1 - p(v_i)) \tag{5.15}$$

where $p(v_i)$ is the packet loss probability for packet v_i in frame f_k, and the frame corruption probability of f_k is defined by the probability that any packet in frame f_k is lost. That is, we consider the whole frame as corrupted as long as part of the frame is lost and the corrupted frame is removed from the reference frame list. This simplifies the reference frame selection algorithm. As a future work, we can further improve our algorithm by performing reference frame selection on partial frames.

We use the frame corruption estimation results to assist the reference frame selection. During the encoding process, we initialize the reference list that consists of previously encoded frames in the same description. Next, we remove the frames with a frame corruption probability greater than a threshold p_{thres} from the list. If all frames are removed from the reference list, we check the previously encoded frames in the other description and add frames with $p(f_k) < p_{\text{thres}}$ to the list. The current frame is encoded using the reference frames in the list and transmitted over the networks. The process is shown in Procedure 1. By not using the possible damaged frame as reference, we expect to reduce error propagation due to packet losses. Moreover, our proposed approach only relies on the standard ad-hoc routing messages and it does not incur any extra overhead.

Despite minimizing frame corruption estimation errors through careful modeling, estimation errors may still occur due to the random feedback delay of routing messages. These unexpected estimation errors can reduce the gains of our design: failing to detect corrupted frames (miss-detections) can lead to error propagation, and incorrectly identifying good frames as corrupted (false-

Procedure 1 Encode a video packet for transmission

```
1: while have video context to send do
2:   Initiate reference list for current frame
3:   for all reference frames in current description do
4:     if the frame corruption probability of the reference is larger than a
       threshold then
5:       Remove the frame from the reference frame list
6:     end if
7:   end for
8:   if no frames are available in the reference list then
9:     Add available frames in the other description to the reference list
10:  end if
11:  Encode a packet of video using selected references and transmit it through
     one of the two paths
12:  if receive a RERR message that implies a link failure then
13:    Estimate the RERR delay and determine the packet loss probability
       for affected packets
14:    if a route is available in the route cache then
15:      Reconstruct a new route from the route cache
16:    else
17:      Initiate the route recovery process
18:      repeat
19:        Mark the packets scheduled to be sent through the broken route
           as lost
20:      until receive a RREP to build a new route
21:    end if
22:    Estimate the frame corruption probability based on the estimated
       packet loss probability
23:  end if
24: end while
```

alarms) can reduce video coding efficiency. To address the estimation errors, we can change the threshold p_{thres} used for frame corruption detection to achieve a flexible tradeoff between the error resilience and coding efficiency. By configuring p_{thres}, we can adapt our design to different scenarios, *e.g.*, we can reduce p_{thres} for the applications that are sensitive to error propagation. In Section 5.6, we study the overall estimation accuracy of our proposed approach under various p_{thres} values.

5.5 Implementation and Simulation Setup

We have simulated the routing-aware MSVC system using the modified JM codec and the Qualnet simulator, and we have examined the performance of the reconstructed video at the receiver under varying network settings. First, we simulated a two-path transport system over a mobile ad-hoc network using the Qualnet simulator. The routing information received at the transmitter is recorded and fedback to the JM encoder. Based on the MAC layer parameters, routing information, and our packet loss estimation model, we estimated the packet loss probability for each packet and used it to guide the reference frame selection during encoding. Then we generated the corrupted video bitstream based on the encoded video sequence and the network simulations. Finally, we decoded the bitstream using our refined error concealment method for MSVC [72]. Details of the network settings, parameters for the estimation model, and video source statistics are described below.

Table 5.1: Simulation Parameters for the QualNet Simulator

Region	500 m× 500 m
Number of nodes	50
Mobility model	Random waypoint model: node speed 0 $\sim$ 10 m/s, pause time 120 s
PHY data rate	5.5 Mbps
Transmission Power	15 dBm
MAC layer protocol	802.11b CSMA/CA
Playout deadline	350 ms

5.5.1 Network Settings for the QualNet Simulator

We use a QualNet simulator to evaluate our routing-aware MSVC over a mobile ad-hoc network. Unless otherwise specified, we choose network parameters as shown in Table 5.1. In the ad-hoc network, nodes are uniformly placed in a $500m \times 500m$ region, where the connectivity of any two nodes is determined by the network topology and the communication range. The movement of each node is characterized by a random waypoint model [81] with parameters shown in the table. A pair of source and destination nodes is randomly chosen to transmit video packets. We use IEEE 802.11b, which employs CSMA/CA as the MAC layer protocol and we implement SMR as the multipath routing protocol. Packets are dropped if they do not reach the destination by the playout deadline of 350 ms.

QualNet uses a wireless communication medium model to simulate the propagation of signals between nodes [82]. This model takes into account propagation delays and signal attenuation due to path loss, fading, and shadowing. In

Table 5.2: Parameters of IEEE 802.11b

Parameter	**Value**
Slot time T_{slot}	20 μs
PHY header	192 bits
MAC header	224 bits
ACK packet	112 bits + PHY header
DIFS time T_{DIFS}	50 μs
SIFS time T_{SIFS}	10 μs
CW_{min}	31
CW_{max}	1023
Retransmission limit	7

our simulation, we choose a two-ray path loss model that considers a line-of-sight path and a reflection from flat earth in the pathloss calculation. We use a Rayleigh fading model to calculate the effect of a propagation path on the signal strength and lognormal shadowing model to calculate the signal attenuation caused by obstruction on a propagation path.

5.5.2 Parameters for The Estimation Model

In Section 5.3, we propose a statistical model to estimate the packet loss probability of each packet transmitted over the networks based on the routing messages and MAC layer parameters. We assume that all the nodes in the network employ the DCF basic access mechanism for packet transmission. The parameters for packet loss estimation are shown in Table 5.2.

5.5.3 Video Source and Performance Metrics

We consider five video sequences "Foreman", "Coastguard", "Mother-daughter", "News", and "Silent", which are all at CIF format with 150 frames at a frame rate of 15 fps. The video sequences are encoded into RTP packets with a packet size of 500 bytes. We generate two descriptions for each video sequence and the bitrate of each video sequence is 400 kbps, which corresponds to a bitrate of 200 kbps for each description. The two descriptions are transmitted through two paths over the network. For each network scenario, each video sequence is sent repeatedly 500 times to generate statistically meaningful quality measures.

We use average PSNR of all frames over all realizations to evaluate the objective video quality of the decoded video sequences. In addition, we use $\text{PSNR}_{r,f}$ introduced in Section 2.3 to evaluate the perceptual video quality for multiple users over the network.

5.6 Performance Evaluation of RA-MSVC

Using the simulation setup described in Section 5.5, we simulated the routing-aware MSVC (RA-MSVC) method with MPT and compared the end-to-end performance with single description coding (SDC) and MSVC with MPT. For SDC and MSVC , we use the same MPT strategy such that even and odd frames are transported through two separate routes. For both RA-MDC and MDC methods, we apply the refined error concealment method proposed in [72] to reconstruct the video. We first study the overall video performance of the

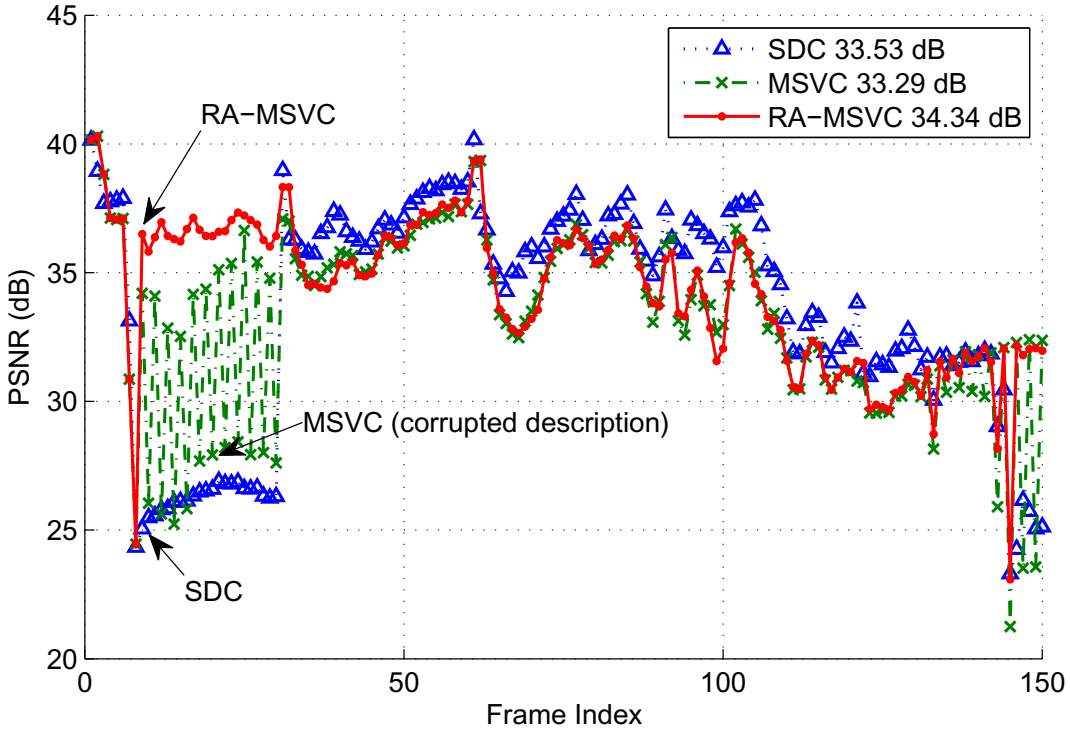


Figure 5.4: PSNRs of Each Frame in One Realization for Foreman Sequence (CIF, 15fps) Coded at 400 kbps. Average PSNRs for SDC, MSVC, and RA-MSVC in This Realization are 33.53 dB, 33.29 dB, and 34.34 dB Respectively

three methods and we examine the estimation accuracy of our proposed model under various network settings.

5.6.1 Overall Performance

First, we examine the case that the transmission power of each node is 15 dBm and the overall packet loss rate in the network is around 4.5%. We show the PSNR values of each frame in one realization in Fig. 5.4. When the packets are transmitted successfully, SDC achieves slightly higher PSNR than MSVC and RA-MSVC because the coding efficiency of MSVC and RA-MSVC decreases

Table 5.3: Average PSNR for Coded Foreman Sequence at 400 kbps Without and With Transmission Losses

PSNR (dB)	SDC	MSVC	RA-MSVC
Without losses	35.77	34.56	34.50
With losses (p=4.5%)	32.20	32.45	33.51

due to the decreased correlation between adjacent frames. When packet loss happens, the PSNR value of the corrupted frame drops and the errors propagate to all subsequent frames of SDC (dashed line with triangle marker) until an I-frame is received. For MSVC (dash-dot line with cross marker), the errors only propagate in the description on the broken route and the PSNR oscillates as shown in Fig. 5.4. Meanwhile, the RERR packets indicate the packet losses in the network fairly accurately and our proposed RA-MSVC (solid line with point marker) method can effectively stop the error propagation in the subsequent frames.

Then we look at the PSNR performance of the three methods without and with transmission losses in Table 5.3. We see that for the coded Foreman sequence without transmission losses, SDC achieves highest PSNR under the same bitrate, while MDC has a PSNR slightly higher than RA-MDC. On the other hand, RA-MDC achieves the highest average PSNR in the presence of moderate transmission losses. The results show that both MDC and RA-MDC trade coding efficiency for the reconstructed video quality under transmission losses, while RA-MDC provides a better tradeoff between coding efficiency and error resilience. Based on the frame loss estimation in RA-MDC, fewer frames are used as reference for RA-MDC, which leads to a 0.06 dB lower PSNR than MDC when there is no transmission loss. However, the RA-MDC achieves 1

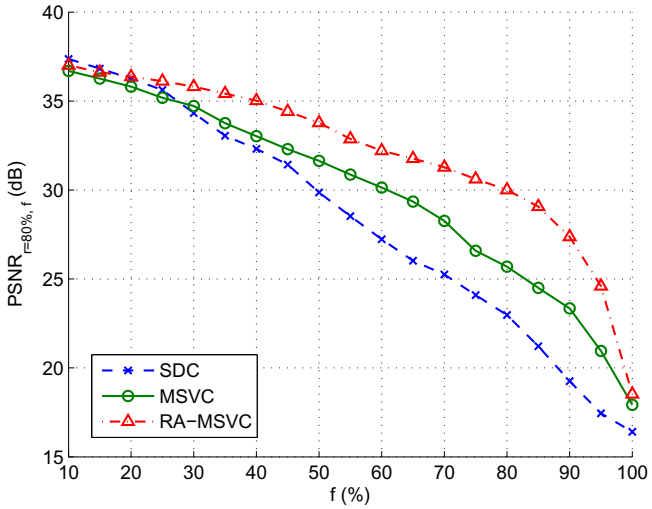
dB gain in PSNR under transmission losses, since it can effectively stop error propagation by not using corrupted frames as reference.

As discussed in Section 2.3, average PSNR among all frames over all realizations does not correlate very well with the perceptual video quality. Therefore, we present $\text{PSNR}_{r,f}$ [83] to assess better the perceptual video quality of the three methods.

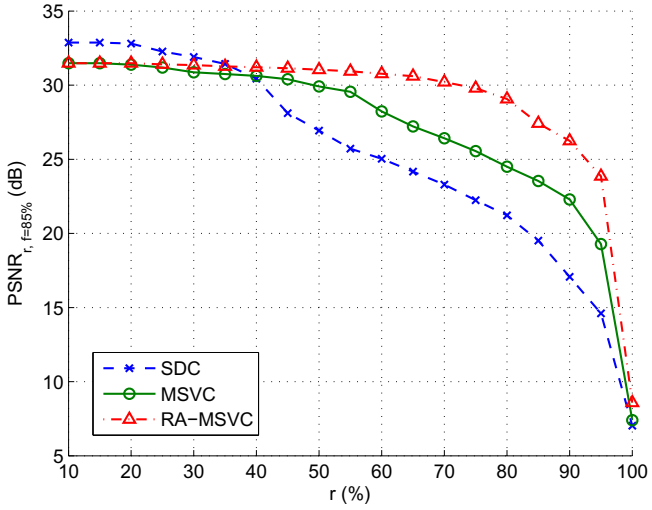
Figure 5.5(a) and 5.5(b) compare the $\text{PSNR}_{r,f}$ results for Foreman sequence with fixed values of r and f, respectively. Figure 5.5(a) shows the $\text{PSNR}_{r,f}$ values for the three coding methods with fixed $r = 80\%$, which indicates the delivered video quality guaranteed for 80% of the users for f percentage of the frames. In Fig. 5.5(a), we see that about 28%, 16%, and 5% of the frames in 80% of the realizations have a PSNR lower than 25 dB for SDC, MSVC, and RA-MSVC, respectively. This shows that RA-MSVC has the fewest bad-quality frames for 80% of the channel uses.

Figure 5.5(b) presents $\text{PSNR}_{r,f}$ for SDC, MSVC, and RA-MSVC with fixed $f = 85\%$. This figure shows that RA-MSVC guarantees a better video quality for most of the realizations compared to the other two methods. For example, RA-MSVC guarantees a PSNR of 29.07 dB for 85% of the frames in 80% of the realizations, while SDC and MSVC can only guarantee a PSNR of 24.50 dB and 21.22 dB for the same values of r and f. This indicates that RA-MSVC provides better video quality for most of the users over the network.

Next, we examine the performance of the three methods under different packet loss rates. In the simulations, we varied the transmission power from 15 dBm to 10 dBm to achieve packet loss rates in the range of 2.2%-12.9%. Figure



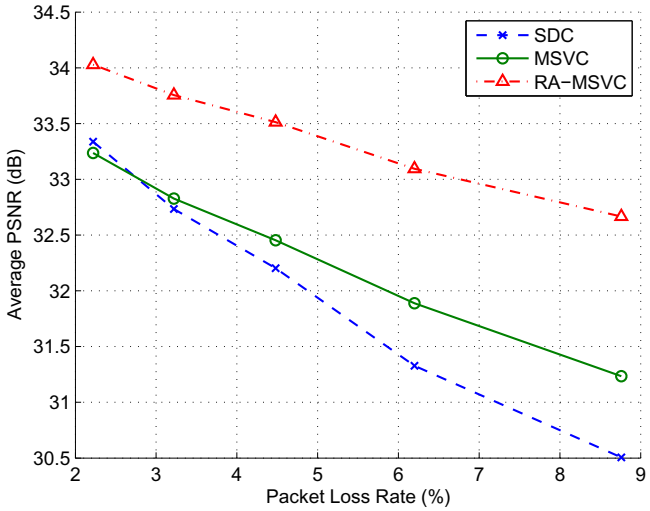
(a) $\text{PSNR}_{r=80\%,f}$



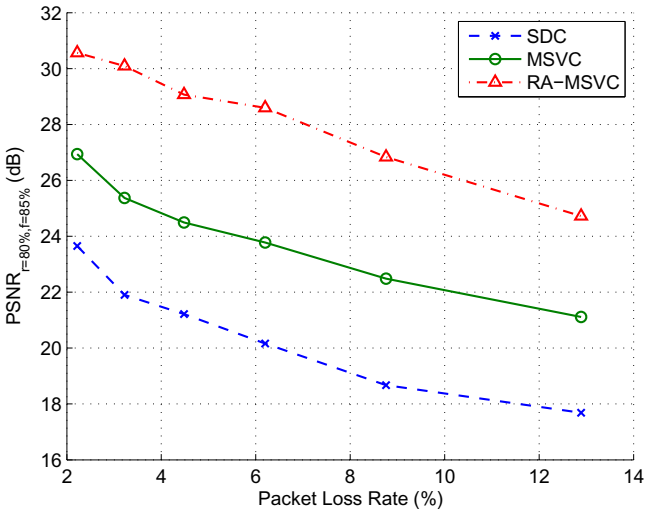
(b) $\text{PSNR}_{r,f=85\%}$

Figure 5.5: Comparing $\text{PSNR}_{r,f}$ of SDC, MSVC, and RA-MSVC, Foreman Sequence (CIF, 15fps) at 400 kbps, Packet Loss Rate 4.5%. Average PSNRs of SDC, MSVC, and RA-MSVC are 32.20 dB, 32.45 dB, and 33.51 dB respectively

5.6(a) presents the average PSNR under different packet loss rates. We see that the gains in PSNR provided by RA-MSVC increase as the packet loss rate



(a) Average PSNR vs Packet Loss Rate



(b) $\text{PSNR}_{r=80\%,f=85\%}$ vs Packet Loss Rate

Figure 5.6: Performance under Different Packet Loss Rates for Foreman Sequence (CIF, 15fps) at 400 kbps. Transmission Power Varies from 10 dBm to 15 dBm to Achieve Different Packet Loss Rates.

increases. This is because the gain of error resilience for RA-MSVC overcomes the reduction in coding efficiency. In Fig. 5.6(b), we present $\text{PNSR}_{r,f}$ under

Table 5.4: Performance for Different Video Sequences under Different Packet Loss Rate

Sequence	packet loss rate 4.5%					
	SDC	MSVC	RA-MSVC	SDC	MSVC	RA-MSVC
	PSNR			$PSNR_{r=80\%,f=85\%}$		
Coastguard	28.30	28.38	29.03	21.42	23.77	27.76
Foreman	32.20	32.45	33.51	21.22	24.49	29.07
Mother-daughter	39.40	39.86	40.63	27.86	34.20	38.66
News	37.24	37.50	38.53	22.47	29.63	36.06
Silent	35.13	35.20	35.89	24.93	29.66	33.85
	packet loss rate 8.8%					
	PSNR			$PSNR_{r=80\%,f=85\%}$		
Coastguard	27.16	27.62	28.52	18.79	22.30	25.16
Foreman	30.51	31.23	32.67	18.67	22.49	26.84
Mother-daughter	37.91	38.91	39.99	24.29	32.00	34.28
News	35.43	36.38	37.72	20.42	27.89	31.41
Silent	33.61	34.32	35.24	20.32	26.59	29.32

different packet loss rates with $r = 80\%$ and $f = 85\%$. The results show that RA-MSVC outperforms MSVC by about 5 dB and outperforms SDC by about 8 dB, which indicates that RA-MSVC provides better video quality for most of the users under various packet loss rates.

Finally, we present the performance of the three methods for five different video sequences under two typical packet loss rates in Table 5.4. These results show that RA-MSVC achieves gains in PSNR in the range of 0.7-2.3 dB compared to SDC, and gains in PSNR in the range of 0.7-1.4 dB compared to MSVC. Furthermore, RA-MSVC increases $PSNR_{r=80\%,f=85\%}$ by up to 13.6 dB as compared to SDC and by up to 6.4 dB as compared to MSVC.

5.6.2 Model Estimation Accuracy

We showed that RA-MSVC improves both objective and perceptual video quality of delivered videos in Section 5.6.1. This indicates that our proposed method can accurately estimate the frame corruption based on the routing messages and network parameters. Now we examine the performance of our estimation process under various network settings to verify its robustness.

We can consider the frame corruption estimation problem as a binary classification problem, in which we try to determine whether a frame is corrupted or not based on the routing information and network conditions. Therefore, we can run a binary hypothesis test to evaluate the performance of our frame corruption estimation. There are two hypotheses: H_0 corresponds to the situation that a frame is correctly received; H_1 corresponds to the situation that a frame is corrupted. Based on our estimation model, we have our estimation outcomes A_0 and A_1, where A_0 means we estimate the frame to be correctly received and A_1 means we treat the frame as corrupted. Then we can use two error probabilities to measure the accuracy of our estimation. $P_{\text{FA}} = P(A_1|H_0)$ is referred to as a false alarm, which corresponds to the probability of detecting a corrupted frame when the frame is actually correctly received. $P_{\text{MISS}} = P(A_0|H_1)$ is referred to as a miss-detection, which corresponds to the probability of detecting a correctly received frame when the frame is actually corrupted.

First, we plot a receiver operating curve (ROC) [84] to represent the possible values of P_{FA} and P_{MISS} under various p_{thres} in the range of [0, 1] in Fig. 5.7. This figure is generated using the default network settings presented in Section 5.5 with transmission power of 15 dBm. In this ROC space, the (0, 0) point

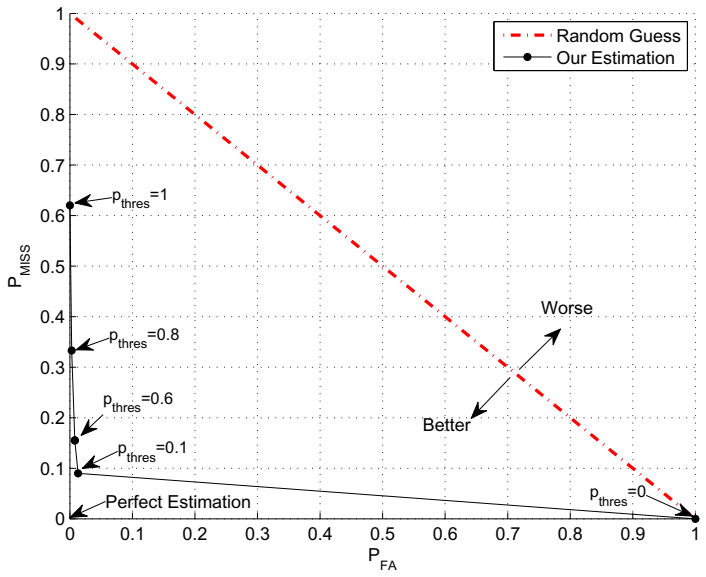


Figure 5.7: ROC Curve for $p_{\text{thres}} \in [0, 1]$. Varying p_{thres} Provides a Trade-off Between P_{FA} and P_{MISS}.

represents perfect estimation and the diagonal line denotes a completely random guess. Therefore, the closer the point is to the lower left corner, the better the overall accuracy of the estimation. In Fig. 5.7, we see that the ROC curve of our estimation is close to the lower left corner for the p_{thres} values in the range of [0.1, 0.6], where we achieve a very low probability of a false classification while maintaining a fairly low probability of missing a corrupted frame. This suggests that by choosing a p_{thres} value in that range, our estimation method yields fairly good performance.

Next, we examine the accuracy of our estimation method under different network settings. Here we choose $p_{\text{thres}} = 0.5$. Table 5.5 presents P_{FA} and P_{MISS} values under different transmission powers. We see that the false alarm probability is constantly low under different transmission powers, which leads

Table 5.5: P_{FA}, P_{MISS} under Different Transmission Powers

Transmission Power	P_{FA}	P_{MISS}
15 dBm	0.008	0.155
14 dBm	0.010	0.195
13 dBm	0.012	0.229
12 dBm	0.014	0.267
11 dBm	0.016	0.315
10 dBm	0.019	0.372

Table 5.6: P_{FA} and P_{MISS} under Different Number of Nodes

Number of Nodes	P_{FA}	P_{MISS}
20	0.008	0.144
40	0.008	0.164
50	0.008	0.155
60	0.008	0.169
80	0.008	0.164

to negligible unnecessary reduction in coding efficiency. The miss-detection probability indicates that our proposed estimation method can detect most of the corrupted frames. We notice that the performance of the estimation becomes worse as the transmission power reduces. This is because when the transmission power decreases, the network connectivity becomes worse. The loss probability of routing messages may increase, which leads to more miss-detections.

We also investigate the impact of the number of nodes on our estimation accuracy. As shown in Table 5.6, varying the number of nodes has little impact on the false alarm probability and miss-detection probability. Similarly, our

experiments show that the estimation accuracy is insensitive to the network size, transmission bitrate, and number of retransmissions.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

In this dissertation, we have designed mechanisms based on MSVC with path diversity to provide error resilient video transmission over mobile wireless networks. We have designed a MB-based error concealment method for MSVC at the decoder side to improve the reconstructed video quality. At the encoder, we applied the rate-distortion optimized mode selection that considers packet loss conditions of the network during the encoding process. In addition, we proposed to use routing feedback to assist reference frame selection and alleviate error propagation.

In Chapter 3, we presented refined error concealment methods for MSVC to improve the error resilience for video communications over wireless ad-hoc networks. The refined intra MB concealment provides better concealment for the MBs in intra frames by using the temporal correlation between adjacent intra frames in two descriptions. The refined inter MB concealment achieves im-

provement from the additional reference list used for motion-compensated concealment. Compared to SDC and the original MSVC, the MSVC_REC method is shown to obtain performance gains over the wireless ad-hoc networks for a wide range of different burst loss rates and random loss rates

In Chapter 4 we presented an error resilient video coding method that enhances the robustness of video to both random loss and burst losses over wireless networks. The method estimates the end-to-end distortion under the specific random and burst loss rates, and applies RD-based mode selection to select the optimal coding mode. The method is applied for single description video coding and multiple description video coding. For single description video coding, we calculate the reconstructed pixel value caused by random loss and burst losses, which results in a more accurate estimation of distortion. The accuracy of the estimation enhances its error robustness over lossy networks. For multiple description video coding, we estimate the distortion for MSVC and optimally select the coding mode in both descriptions. Compared to MSVC, our approach alleviates error propagation due to random loss in the two descriptions of MSVC and achieves better performance than MSVC under both random packet loss and burst losses. Note that the complexity of our approach, which is only incurred at the encoder, is comparable to the ROPE approach.

In Chapter 5, we presented a routing-aware MDC approach with multipath transport to enhance the error robustness of video transmission over wireless ad-hoc networks. We establish a model to estimate the packet loss probability of each packet based on routing messages and network parameters. Then we use the estimated packet loss probability to select the proper reference frames

for MDC in order to reduce error propagation. Our proposed method does not require any additional feedback channel or extra overhead while it nicely captures the potential frame corruption during transmission. We examine our proposed RA-MDC method using a modified JM coder and the QualNet simulator. The simulation results show that our method achieves up to 2.3 dB gains in PSNR for different video sequences under different network conditions. Using $\text{PSNR}_{r,f}$ as a multiuser perceptual quality measure, the results also indicate that RA-MDC guarantees better perceptual video quality for multiple users. In addition, we show that our proposed method has good estimation accuracy under various network settings, which leads to the improvement of the delivered video.

6.2 Future Work

In the following subsections, we introduce some future research directions based on the work in this dissertation.

6.2.1 Enhanced Error Concealment for Multiple Description Coding

Shannon has shown that any redundancy in the source will help to combat noise if it is properly utilized at the receiver [85]. In our proposed system in Fig. 5.1, routing-aware MDC is applied to increase the error robustness of the video. Since MDC introduces certain redundancy among descriptions, it is

important to develop an appropriate error concealment method to obtain better recovery.

In [72], we have developed MB-level error concealment methods for both MDC and RA-MDC by exploring correlation across different descriptions. Furthermore, a frame loss concealment method for RA-MDC may be explored to recover the loss of a whole frame. When frames in one description are dropped during transmission, an adequate frame loss concealment method will reduce the PSNR fluctuation among descriptions while improve perceptual video quality for users.

6.2.2 Routing-aware Motion and Mode Selection

Our prior work uses routing information to guide the selection of reference frames, which effectively alleviates error propagation. Basically, the frame with a high frame corruption probability is not used as a reference frame. However, the packet drops during transmission may not correspond to the whole frame loss, which means the uncorrupted part of the frame could still be used for motion compensation. Instead of removing the whole frame from the reference frame list, a routing-aware motion selection and mode selection method can be designed to adapt to the dynamic networks.

In a typical rate-distortion optimization framework for a video codec, the encoder first determines the reference frame and the associated motion vector for a specific coding mode in a MB by minimizing the distortion subject to a rate constraint. Then the encoder chooses the optimal coding mode for each MB

using a similar rate distortion optimization technique. Usually, the distortion calculated is the compression distortion.

With routing feedback from the network layer, we are able to estimate packet loss probabilities of each transmitted packet [86]. To better select the proper reference frame, motion vector, and coding mode for each MB, the distortion due to compression and the distortion introduced by the possible corruption of a motion compensated block in the reference frame need to be considered. The predicted pixel value for the chosen *ref* and MV can be estimated as

$$\tilde{f}_{ref}(\Delta x, \Delta y) = (1 - p_i) \cdot \hat{f}_{ref}(\Delta x, \Delta y) + p_i \cdot \bar{f}_{ref}(\Delta x, \Delta y) \tag{6.1}$$

where $(\Delta x, \Delta y)$ is the coordinates of the motion compensated pixel, $\hat{f}_{ref}(\Delta x, \Delta y)$ is the correctly reconstructed pixel value of $(\Delta x, \Delta y)$ in frame *ref*, $\bar{f}_{ref}(\Delta x, \Delta y)$ is the concealed pixel value when packet i is lost and p_i is the estimated packet loss probability of packet i (using the estimation model proposed in [86]). The overall distortion will be estimated based on the value of $\tilde{f}_{ref}(\Delta x, \Delta y)$ and used to decide the optimal motion vector, reference frame and coding mode for each MB.

6.2.3 Transmission Strategies and Rate Allocation

In light loaded wireless ad hoc networks, packet losses in the networks are mostly caused by node mobility but not by network congestions. We have shown the effectiveness of RA-MDC under such light loaded networks (one single low-bitrate video flow without any background traffic) [87]. Nowadays, wireless ad hoc networks are becoming increasingly congested due to the quick increase

of wireless demands. In such networks, wireless congestion can contribute to a large portion of packet losses as well as route failures. Therefore, it is important to consider heavily loaded networks.

The transmission strategies and rate allocation for RA-MDC need to be investigated in a heavily loaded network with various types of background traffic. When a route is lost, the transmitter can either drop the packets on the lost route or continue to transmit all the packets through the remaining good route. The latter approach reduces packet drops, but at a cost of introducing higher loads on the good route. In a heavily loaded network, such a load increase on one route may lead to network congestion and the increase of packet delay. The video coding rate would be adapted to the network conditions to alleviate the load of the networks.

Bibliography

[1] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 966–976, Jun. 2000.

[2] Cisco visual networking index: Global mobile data traffic forecast update, 2009-2014. [Online]. Available: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.pdf

[3] A. Nafaa, T. Taleb, and L. Murphy, "Forward error correction strategies for media streaming over wireless networks," *IEEE Communications Magazine*, vol. 46, no. 1, pp. 72–79, Jan. 2008.

[4] Y. Zhang, W. Gao, Y. Lu, Q. Huang, and D. Zhao, "Joint source-channel rate-distortion optimization for H.264 video coding over error-prone networks," *IEEE Transactions on Multimedia*, vol. 9, no. 3, pp. 445–454, Apr. 2007.

[5] Y. Wang, A. R. Reibman, and S. Lin, "Multiple description coding for video delivery," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 57–70, Jan. 2005.

[6] H.-y. Luo, Z.-l. Gan, and X.-c. Zhu, "Content-adaptive interpolation for spatial error concealment," in *Second International Conference on Information and Computing Science*, vol. 1, 2009, pp. 265–268.

[7] J. Wu, X. Liu, and K.-Y. Yoo, "A temporal error concealment method for H.264/AVC using motion vector recovery," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 4, pp. 1880–1885, Nov. 2008.

[8] Y. Chen, Y. Hu, O. C. Au, H. Li, and C. W. Chen, "Video error concealment using spatio-temporal boundary matching and partial differential

equation," *IEEE Transactions on Multimedia*, vol. 10, no. 1, pp. 2–15, Jan. 2008.

[9] M. hsuan Lu, P. Steenkiste, and T. Chen, "Robust wireless video streaming using hybrid spatial/temporal retransmission," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 3, pp. 476–487, Apr. 2010.

[10] T. Wiegand, N. Farber, K. Stuhlmuller, and B. Girod, "Error-resilient video transmission using long-term memory motion-compensated prediction," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 1050–1062, Jun 2000.

[11] M. Budagavi and J. D. Gibson, "Multiframe video coding for improved performance over wireless channels," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 252 –265, Feb. 2001.

[12] V. K. Goyal, "Multiple description coding: compression meets the network," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 74–93, Sep. 2001.

[13] J. G. Apostolopoulos, "Reliable video communication over lossy packet networks using multiple state encoding and path diversity," in *SPIE Proceedings on Visual Communications and Image Processing*, vol. 4310, no. 1, 2001, pp. 392–409.

[14] N. Franchi, M. Fumagalli, R. Lancini, and S. Tubaro, "Multiple description video coding for scalable and robust transmission over IP," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 3, pp. 321–334, Mar. 2005.

[15] A. Reibman, H. Jafarkhani, Y. Wang, and M. Orchard, "Multiple description video using rate-distortion splitting," in *Proceedings of International Conference on Image Processing*, vol. 1, 2001, pp. 978–981.

[16] V. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 821–834, May 1993.

[17] S. Dumitrescu and X. Wu, "On properties of locally optimal multiple description scalar quantizers with convex cells," *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5591–5606, Dec. 2009.

[18] Y. Wang, M. T. Orchard, V. Vaishampayan, and A. R. Reibman, "Multiple description coding using pairwise correlating transforms," *IEEE Transactions on Image Processing*, vol. 10, no. 3, pp. 351–366, Mar. 2001.

[19] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, July 2003.

[20] J. Hu, S. Choudhury, and J. D. Gibson, "$PSNR_{r,f}$: Assessment of delivered AVC/H. 264 video quality over 802.11a WLANs with multipath fading," in *Proceedings of the First Multimedia Communications Workshop*, 2006.

[21] Z. Wang, H. R. Sheikh, and A. C. Bovik, "Objective video quality assessment," in *The Handbook of Video Databases: Design and Applications*. CRC Press, 2003, pp. 1041–1078.

[22] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: a review," *Proceedings of the IEEE*, vol. 86, no. 5, pp. 974–997, May 1998.

[23] S. Aign and K. Fazel, "Temporal and spatial error concealment techniques for hierarchical MPEG-2 video codec," in *IEEE International Conference on Communications*, vol. 3, 1995, pp. 1778 –1783.

[24] W. Kwok and H. Sun, "Multi-directional interpolation for spatial error concealment," *IEEE Transactions on Consumer Electronics*, vol. 39, no. 3, pp. 455–460, 8-10 1993.

[25] W. Zeng and B. Liu, "Geometric-structure-based error concealment with novel applications in block-based low-bit-rate coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 4, pp. 648–665, Jun. 1999.

[26] R. Zhang, Y. Zhou, and X. Huang, "Content-adaptive spatial error concealment for video communication," *IEEE Transactions on Consumer Electronics*, vol. 50, no. 1, pp. 335–341, Feb. 2004.

[27] W.-Y. Kung, C.-S. Kim, and C.-C. Kuo, "Spatial and temporal error concealment techniques for video transmission over noisy channels," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 7, pp. 789–803, July 2006.

[28] S. H. Hyun, S. S. Kim, B. C. Kim, I. K. Eom, and Y. S. Kim, "Efficient directional interpolation for block recovery using difference values of border pixels," in *Congress on Image and Signal Processing*, vol. 3, 2008, pp. 565–568.

[29] S. Belfiore, L. Crisa, M. Grangetto, E. Magli, and G. Olmo, "Robust and edge-preserving video error concealment by coarse-to-fine block replenishment," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, 2002, pp. IV–3281–3284.

[30] S. S. Hemami and T. H.-Y. Meng, "Transform coded image reconstruction exploiting interblock correlation," *IEEE Transactions on Image Processing*, vol. 4, no. 7, pp. 1023–1027, Jul. 1995.

[31] H. Y. Lee, I. K. Eom, and Y. S. Kim, "Error concealment using directional coefficient mask and difference of DC," in *30th Annual Conference of IEEE Industrial Electronics Society*, vol. 3, 2004, pp. 2086–2091.

[32] Y. Wang, Q.-F. Zhu, and L. Shaw, "Maximally smooth image recovery in transform coding," *IEEE Transactions on Communications*, vol. 41, no. 10, pp. 1544–1551, Oct. 1993.

[33] J. W. Park, J. W. Kim, and S. U. Lee, "DCT coefficients recovery-based error concealment technique and its application to the MPEG-2 bit stream error," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 6, pp. 845–854, Dec. 1997.

[34] X. Lee, Y.-Q. Zhang, and A. Leon-Garcia, "Information loss recovery for block-based image coding techniques-a fuzzy logic approach," *IEEE Transactions on Image Processing*, vol. 4, no. 3, pp. 259–273, Mar. 1995.

[35] H. Sun and W. Kwok, "Concealment of damaged block transform coded images using projections onto convex sets," *IEEE Transactions on Image Processing*, vol. 4, no. 4, pp. 470–477, Apr. 1995.

[36] P. Haskell and D. Messerschmitt, "Resynchronization of motion compensated video affected by ATM cell loss," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, 1992, pp. 545–548.

[37] W. M. Lam, A. R. Reibman, and B. Liu, "Recovery of lost or erroneously received motion vectors," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, 1993, pp. 417–420.

[38] J. Zhang, J. F. Arnold, and M. R. Frater, "A cell-loss concealment technique for MPEG-2 coded video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 4, pp. 659–665, Jun. 2000.

[39] S. Tsekeridou and I. Pitas, "MPEG-2 error concealment based on block-matching principles," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 4, pp. 646 –658, Jun. 2000.

[40] P. Salama, N. B. Shroff, and E. J. Delp, "Error concealment in MPEG video streams over ATM networks," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 1129–1144, Jun. 2000.

[41] Z.-H. Zhou, S.-L. Xie, and Z.-L. Xu, "Efficient adaptive MRF-MAP error concealment of video sequences," in *Proceedings of International Conference on Machine Learning and Cybernetics*, vol. 9, 2005, pp. 5507–5511.

[42] S. Shirani, F. Kossentini, and R. Ward, "A concealment method for video communications in an error-prone environment," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 1122–1128, Jun. 2000.

[43] L. Atzori, F. De Natale, and C. Perra, "A spatio-temporal concealment technique using boundary matching algorithm and mesh-based warping (BMA-MBW)," *IEEE Transactions on Multimedia*, vol. 3, no. 3, pp. 326–338, Sep. 2001.

[44] Y.-C. Lee and Y. Altunbasak, "A collaborative multiple description transform coding and statistical error concealment method for error resilient video streaming over noisy channels," in *Proceedings of IEEE International Conference onAcoustics, Speech, and Signal Processing*, vol. 2, 2002, pp. 2077–2080.

[45] D. Wang, N. Canagarajah, D. Agrafiotis, and D. Bull, "Error concealment for slice group based multiple description video coding," in *IEEE International Conference on Image Processing*, vol. 1, 2005, pp. I–769–772.

[46] D. Wang, N. Canagarajah, and D. Bull, "Slice group based multiple description video coding with three motion compensation loops," in *IEEE International Symposium on Circuits and Systems*, 2005, pp. 960–963.

[47] Y. Lu, R. Zhou, H. Cui, and K. Tang, "Bi-directional entire frame recovery in MDC video streaming," in *IEEE International Symposium on Communications and Information Technology*, vol. 2, 2005, pp. 1058–1061.

[48] M. Ma, O. C. Au, L. Guo, S.-H. G. Chan, and P. H. W. Wong, "Error concealment for frame losses in MDC," *IEEE Transactions on Multimedia*, vol. 10, no. 8, pp. 1638–1647, Dec. 2008.

[49] K. P. Lim, G. Sullivan, and T. Wiegand, "Text description of joint model reference encoding methods and decoding concealment methods," *Joint Video Team of ISO/IEC MPEG and ITU-T VCEG, JVT-O079*, vol. 6, Apr. 2005.

[50] V. Varsa, M. Hannuksela, and Y. Wang, "Non-normative error concealment algorithms," *ITU-T, VCEG-N62*, Sep. 2001.

[51] S. Wenger, "H.264/AVC over IP," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 645–656, July 2003.

[52] Y. J. Liang, J. G. Apostolopoulos, and B. Girod, "Analysis of packet loss for compressed video: Effect of burst losses and correlation between error frames," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 7, pp. 861–874, July 2008.

[53] Q.-F. Zhu and L. Kerofsky, "Joint source coding, transport processing, and error concealment for H.323-based packet video," in *Visual Communications and Image Processing*, vol. 3653, no. 1. SPIE, 1998, pp. 52–62. [Online]. Available: http://link.aip.org/link/?PSI/3653/52/1

[54] R. O. Hinds, T. N. Pappas, and J. S. Lim, "Joint block-based video source/channel coding for packet-switched networks," in *Visual Communications and Image Processing*, vol. 3309, no. 1. SPIE, 1998, pp. 124–133. [Online]. Available: http://link.aip.org/link/?PSI/3309/124/1

[55] G. Cote and F. Kossentini, "Optimal intra coding of blocks for robust video communication over the internet," in *Signal Processing: Image Communication*, vol. 15, no. 1, 1999, pp. 25–34.

[56] G. Cote, S. Shirani, and F. Kossentini, "Optimal mode selection and synchronization for robust video communications over error-prone networks," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 952–965, Jun. 2000.

[57] T. Stockhammer, D. Kontopodis, and T. Wiegr, "Rate-distortion optimization for JVT/H.26L video coding in packet loss environment," in *Proceedings of International Packet Video Workshop*, 2002.

[58] Y. Eisenberg, F. Zhai, T. N. Pappas, R. Berry, and A. K. Katsaggelos, "VAPOR: variance-aware per-pixel optimal resource allocation," *Image Processing, IEEE Transactions on*, vol. 15, no. 2, pp. 289–299, Feb. 2006.

[59] A. Reibman, "Optimizing multiple description video coders in a packet loss environment," in *Packet Video Workshop*, 2002.

[60] B. A. Heng, J. G. Apostolopoulos, and J. S. Lim, "End-to-end rate-distortion optimized MD mode selection for multiple description video coding," *EURASIP Journal on Applied Signal Processing*, pp. 261–261, Jan. 2006.

[61] Y. Liao and J. D. Gibson, "Rate-distortion based mode selection for video coding over wireless networks with burst losses," in *17th International Packet Video Workshop*, 2009, pp. 1–10.

[62] T. Wiegand, M. Lightstone, D. Mukherjee, T. G. Campbell, and S. K. Mitra, "Rate-distortion optimized mode selection for very low bit rate video coding and the emerging H.263 standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 2, pp. 182–190, Apr. 1996.

[63] S. Lin, S. Mao, Y. Wang, and S. Panwar, "A reference picture selection scheme for video transmission over ad-hoc networks using multiple paths," in *Proceedings of IEEE International Conference on Multimedia and Expo*, 2001, pp. 96–99.

[64] Y. J. Liang, E. Setton, and B. Girod, "Channel-adaptive video streaming using packet path diversity and rate-distortion optimized reference picture selection," in *IEEE Workshop on Multimedia Signal Processing*, Dec. 2002, pp. 420–423.

[65] N. Gogate, D.-M. Chung, S. S. Panwar, and Y. Wang, "Supporting image and video applications in a multihop radio environment using path diversity and multiple description coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 9, pp. 777–792, Sep. 2002.

[66] S. Mao, S. Lin, S. S. Panwar, Y. Wang, and E. Celebi, "Video transport over ad hoc networks: multistream coding with multipath transport," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 10, pp. 1721–1737, Dec. 2003.

[67] A. C. Begen, Y. Altunbasak, and O. Ergun, "Multi-path selection for multiple description encoded video streaming," in *IEEE International Conference on Communications*, vol. 3, 2003, pp. 1583–1589.

[68] S. Mao, Y. T. Hou, X. Cheng, H. D. Sherali, S. F. Midkiff, and Y.-Q. Zhang, "On routing for multiple description video over wireless ad hoc networks," *IEEE Transactions on Multimedia*, vol. 8, no. 5, pp. 1063–1074, Oct. 2006.

[69] S. Kompella, S. Mao, Y. T. Hou, and H. D. Sherali, "Path selection and rate allocation for video streaming in multihop wireless networks," in *Proceedings of IEEE Military Communications Conference*, 2006, pp. 1–7.

[70] S. Murthy, P. Hegde, V. Parameswaran, B. Li, and A. Sen, "Improved path selection algorithms for multipath video streaming in wireless ad-hoc networks," in *Proceedings of IEEE Military Communications Conference*, 2007, pp. 1–7.

[71] L. Zhou, B. Geller, B. Zheng, A. Wei, and J. Cui, "System scheduling for multi-description video streaming over wireless multi-hop networks," *IEEE Transactions on Broadcasting*, vol. 55, no. 4, pp. 731–741, Dec. 2009.

[72] Y. Liao and J. D. Gibson, "Refined error concealment for multiple state video coding over ad hoc networks," in *Proceedings of the 42nd Asilomar Conference on Signals, Systems and Computers*, 2008, pp. 2243–2247.

[73] A. Nasipuri and S. R. Das, "On-demand multipath routing for mobile ad hoc networks," in *Proceedings of Eight International Conference on Computer Communications and Networks*, 1999, pp. 64 –70.

[74] S.-J. Lee and M. Gerla, "Split multipath routing with maximally disjoint paths in ad hoc networks," in *Proceedings of IEEE International Conference on Communications*, vol. 10, 2001, pp. 3201–3205.

[75] M. K. Marina and S. R. Das, "On-demand multipath distance vector routing in ad hoc networks," in *Ninth International Conference on Network Protocols*, Nov. 2001, pp. 14 – 23.

[76] Z. Ye, S. V. Krishnamurthy, and S. K. Tripathi, "A framework for reliable routing in mobile ad hoc networks," in *Twenty-Second Annual Joint Conference of the IEEE Computer and Communications*, vol. 1, 2003, pp. 270 – 280.

[77] D. B. Johnson and D. A. Maltz, "Dynamic source routing in ad hoc wireless networks," in *Mobile Computing*. Springer, 1996, vol. 353, pp. 153–181.

[78] S. Mueller, R. P. Tsang, and D. Ghosal, "Multipath routing in mobile ad hoc networks: Issues and challenges," in *Performance Tools and Applications to Networked Systems*, ser. Lecture Notes in Computer Science. Springer, 2004, vol. 2965, pp. 209–234.

[79] Y. Liao and J. D. Gibson, "Routing-aware multiple description video coding over wireless ad-hoc networks using multiple paths," in *International Conference on Image Processing (ICIP)*, Sep. 26-30, 2010.

[80] "IEEE Standard Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," *IEEE Std 802.11-2007 (Revision of IEEE Std 802.11-1999)*, pp. C1–1184, 2007.

[81] J. Broch, D. A. Maltz, D. B. Johnson, Y. C. Hu, and J. Jetcheva, "A performance comparison of multi-hop wireless ad hoc network routing protocols," in *Proceedings of the 4th annual ACM/IEEE international conference on Mobile computing and networking*, 1998, pp. 85–97.

[82] *QualNet 4.5.1 Programmers Guide*, Scalable Network Technologies, July 2008.

[83] J. Hu, S. Choudhury, and J. D. Gibson, "Video capacity of WLANs with a multiuser perceptual quality constraint," *IEEE Transactions on Multimedia*, vol. 10, no. 8, pp. 1465–1478, Dec. 2008.

[84] R. D. Yates and D. J. Goodman, *Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers*. John Wiley & Sons, 2004.

[85] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, Jul., Oct. 1948.

[86] Y. Liao and J. D. Gibson, "Frame corruption estimation from route messages for video coding over mobile ad hoc networks," in *the 44th Annual Asilomar Conference on Signals, Systems, and Computers*, Nov. 7-10, 2010.

[87] ——, "Routing-aware multiple description video coding over mobile ad-hoc networks," *IEEE Transactions on Multimedia, to appear*, 2011.