Video Communications Over Wireless Ad-Hoc Networks Using Source Coding Diversity and Multiple Paths

Yiting Liao and Jerry D. Gibson
University of California, Santa Barbara
USA

1. Introduction

Providing reliable video communications over wireless ad-hoc networks is becoming increasingly important as these networks become widely employed in military, homeland defense security, and disaster recovery applications. However, wireless ad-hoc networks have a dynamically changing topology that can cause failures of links and nodes, thus resulting in path loss. Additionally, video communications over wireless ad-hoc networks can suffer from noise and fading effects in the channel. Therefore, it is important to provide error resilience for reliable video communications over such an error-prone network.

A number of solutions have been proposed for this problem, including source coding diversity and multipath routing. Source coding diversity methods such as multiple description coding (MDC) have proven to be effective for robust video communications, especially when combined with network path diversity (Gogate et al., 2002; Mao et al., 2003; Apostolopoulos & Trott, 2004). We investigate new MDC methods combined with path diversity to enhance the error resilience of video communications over wireless ad hoc networks. The basic idea of MDC is to encode the video sequence into several descriptions for transmission over multiple paths. Each description can be independently decoded and combined with the other descriptions to provide an acceptable video quality. When more descriptions are received for reconstruction, higher video quality can be achieved. As long as all descriptions are not lost simultaneously, somewhat 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 path diversity, which reduces the possibility of simultaneous loss of different descriptions and enables load balancing in networks.

Many MDC algorithms have been proposed (Goyal, 2001) and they can be divided into three categories: subsampling algorithms in the temporal (Apostolopoulos, 2001), spatial (Franchi et al., 2005) or frequency domain (Reibman et al., 2001), multiple description quantization algorithms (Vaishampayan, 1993; Dumitrescu & Wu, 2009), and multiple description transform coding (Wang et al., 2001). Wang, Reibman & Lin (2005) provides a good review for MDC algorithms.

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 any corresponding correlation is used to recover a lost description. One of the most popular MDC methods is multiple state video coding (MSVC) (Apostolopoulos, 2001), which temporally downsamples the video sequence and uses the correlation between adjacent frames in two descriptions to recover from frame loss. More details about MSVC are discussed in Section 2.

We use a rate-distortion optimized mode selection framework to estimate the end-to-end distortion for MSVC by considering the network conditions and multiple state recovery. The estimated end-to-end distortion is used to select the optimal coding mode to reduce error propagation due to packet losses. At the decoder, we investigate a refined error concealment method that uses correlation between different descriptions to better reconstruct the corrupted frames. This method provides better concealment for intra macroblocks (MBs) by using temporal correlation between adjacent intra frames and improves inter MB concealment by using additional reference frames for motion-compensated concealment. We present the performance of these methods over a wireless ad-hoc network with random and burst losses. The results show that the proposed method achieves improvements in objective video quality for a wide range of different burst and random packet loss rates. In addition, we use a multiuser perceptual video quality indicator to capture the distribution of distorted frames across all video frames and multiple channel uses. When combined with MSVC and path diversity techniques, our proposed methods provide better perceptual video experience for multiple network uses.

2. Multiple description video coding with path diversity

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.

Among the many proposed MDC algorithms (Wang, Reibman & Lin, 2005), multiple state video coding (MSVC) proposed by Apostolopoulos (2001) is a very popular method since it is easy to implement and compatible with different video standards. In MSVC, the system includes a multiple state video encoder/decoder and a path diversity transmission system as shown in Fig. 1.

Description 1: $I_1 \rightarrow P_3 \rightarrow P_5 \dots I_{n+1} \rightarrow P_{n+3} \rightarrow P_{n+5} \dots$

Description 2: $I_2 \rightarrow P_4 \rightarrow P_6 \dots I_n \rightarrow P_{n+2} \rightarrow P_{n+4} \dots$
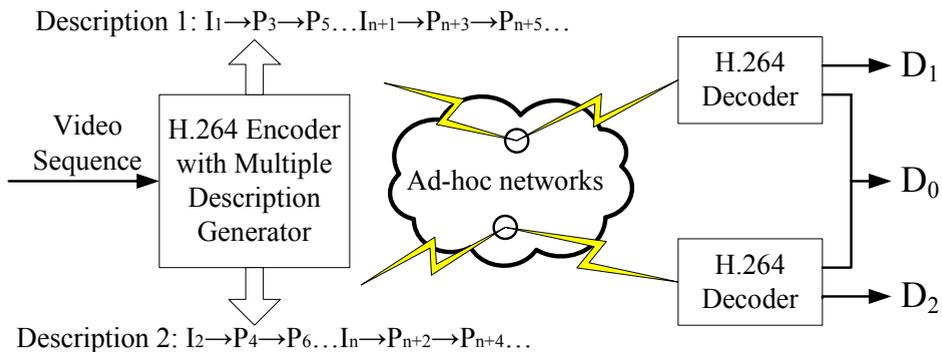
Fig. 1. MSVC system architecture

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 (Wiegand et al., 2003) and transmitted over the networks in two different paths. At the decoder, they are decoded and interleaved to get the reconstructed video sequence.

When one description experiences packet loss, the information in the other description can be used to improve the recovery of the corrupted video segment. This is referred as multiple state recovery (Apostolopoulos, 2001). In Liao & Gibson (2008), the performance of MSVC is further improved by applying refined error concealment methods on a MB basis. For MSVC, the even and odd frames are transmitted in different paths, so burst losses in one description can be well concealed by frames in the other description and cause less damage to the reconstructed video than for single description coding (SDC), where burst losses may result in the loss of consecutive frames. However, random packet loss may cost error propagation in both descriptions and we try to alleviate the error propagation by applying optimal mode selection for MSVC.

3. MSVC with refined error concealment

When video is transmitted over wireless networks, a typical maximum transfer unit (MTU) size is around 100 bytes (Wenger, 2003), which means each frame consists of more than one packet. Therefore, a packet loss only causes some MB losses in a frame. In Apostolopoulos (2001), it is assumed that every packet loss leads to one entire frame loss and the state recovery methods introduced are on a frame basis. Therefore, we propose the refined error concealment methods on a MB basis to enhance the reconstructed video quality for MSVC (Liao & Gibson, 2008). We refer to the approach as multiple state video coding with refined error concealment (MSVC_REC).

3.1 Related work

Error concealment techniques, which have been well developed for decades (Wang & Zhu, 1998), 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 & Fazel (1995) proposed to interpolate lost pixel values from the boundary pixels of the four neighboring MBs. In addition, Kwok & Sun (1993), Zeng & Liu (1999), Zhang et al. (2004), Kung et al. (2006), Hyun et al. (2008), Luo et al. (2009) 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) (Belfiore et al., 2002) 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 (Hemami & Meng, 1995; Lee et al., 2004), performing optimization based on a smoothness constraint (Wang et al., 1993; Park et al., 1997), using the fuzzy logic approach to recover the high-frequency components (Lee et al., 1995), or using an iterative procedure called "projections onto convex sets" (POCS) (Sun & Kwok, 1995).

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 & Messerschmitt (1992) discussed the use of zero MV, the MV of the co-located MB in the previous frame, and the average or median MV of spatially adjacent MBs for the lost MB. The boundary matching algorithm (BMA) (Lam et al., 1993) is proposed to select the best MV among a set of candidate MVs. Zhang, Arnold & Frater (2000), Tsekeridou & Pitas (2000), Kung et al. (2006), Wu et al. (2008) presented different block matching techniques that estimate the MV based on the set of surrounding MBs of the lost MB. Salama et al. (2000) 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. This method is further improved by using an adaptive Huber function in an MRF model (Zhou et al., 2005).

More recently, hybrid algorithms have been proposed to obtain better recovery. They are effective but generally introduce more complexity. Shirani et al. (2000) 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. Atzori et al. (2001) proposed a concealment method which replaces the lost MB using BMA and applies a mesh-based warping (MBW) to reduce the artifacts. In Chen et al. (2008), 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 traditional methods is constrained by the fact that 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 & Altunbasak (2002) adopted a MAP estimation approach to conceal the corrupted description in multiple description transform coding (Wang et al., 2001) and Wang, Canagarajah, Agrafiotis & Bull (2005) proposed error concealment method for a three-loop slice group MDC approach (Wang, Canagarajah & Bull, 2005). In Apostolopoulos (2001), Lu et al. (2005), and Ma et al. (2008), different concealment methods have been 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 sub-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 (Lim et al., 2005). As shown in Fig 2, 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

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

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. 2.

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
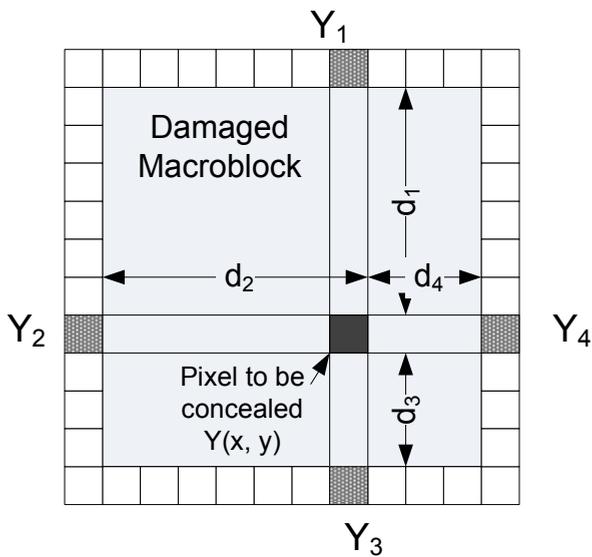
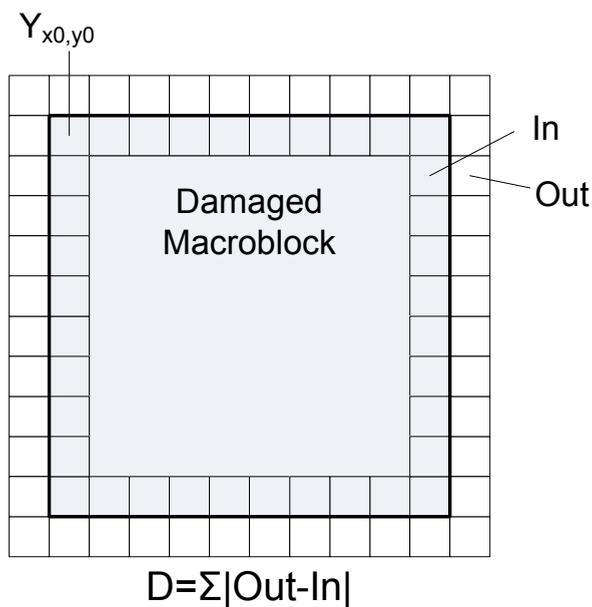Fig. 2. Intra MB Concealment in H.264
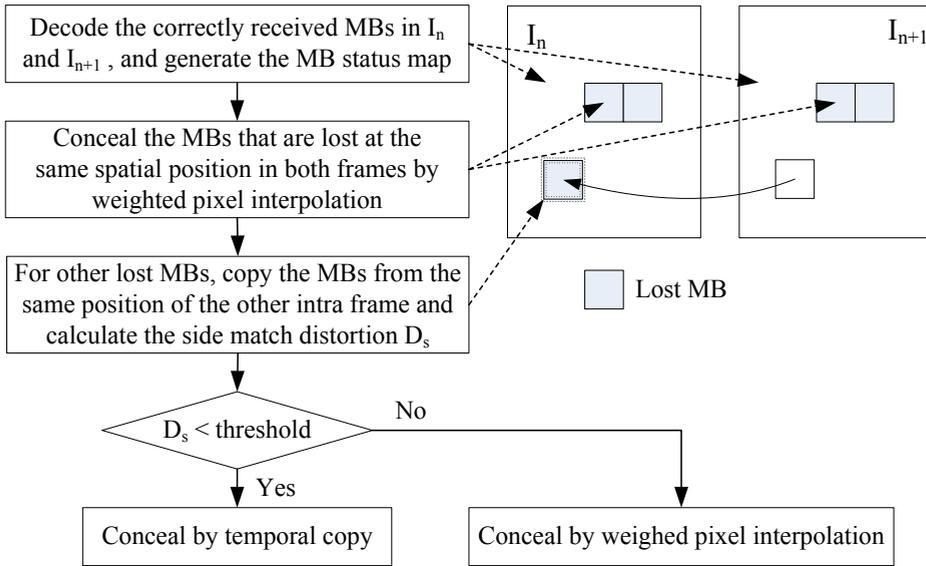

Fig. 3. Side Match Distortion

Fig. 4. Error concealment in intra frame for MSVC

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. 1. 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. 4. 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. 2 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 (Lam et al., 1993) based on the correctly received neighbor MBs. As shown in Fig. 3, the side match distortion D_{SM} is the sum of absolute luminance differences between the concealed MB and neighboring MBs at the boundary,

$$
\begin{aligned}
D_{SM} \quad = \quad & \sum_{i=0}^{15} |Y_{x_0+i,y_0} - Y_{x_0+i,y_0-1}| + \sum_{i=0}^{15} |Y_{x_0,y_0+i} - Y_{x_0-1,y_0+i}| \\
& + \sum_{i=0}^{15} |Y_{x_0+i,y_0+15} - Y_{x_0+i,y_0+16}| + \sum_{i=0}^{15} |Y_{x_0+15,y_0+i} - Y_{x_0+16,y_0+i}| \quad (2)
\end{aligned}
$$

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.
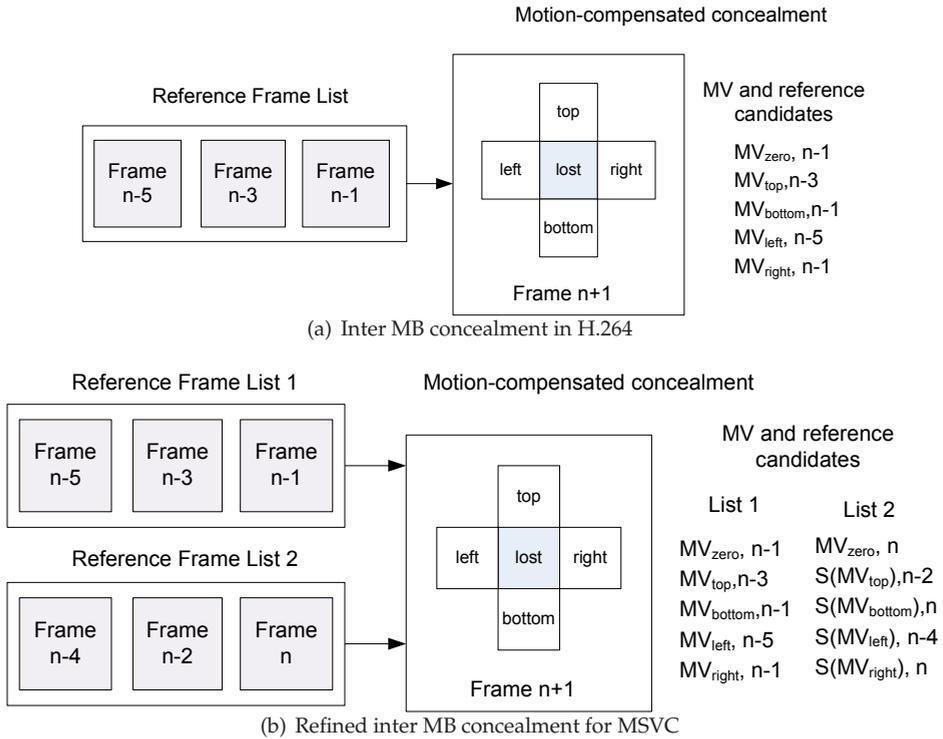
Motion-compensated concealment

Reference Frame List

| Frame n-5 | Frame n-3 | Frame n-1 |

top

left | lost | right

bottom

Frame n+1

MV and reference candidates

MV_{zero}, n-1
MV_{top}, n-3
MV_{bottom}, n-1
MV_{left}, n-5
MV_{right}, n-1

(a) Inter MB concealment in H.264

Reference Frame List 1

| Frame n-5 | Frame n-3 | Frame n-1 |

Reference Frame List 2

| Frame n-4 | Frame n-2 | Frame n |

Motion-compensated concealment

top

left | lost | right

bottom

Frame n+1

MV and reference candidates

List 1

MV_{zero}, n-1
MV_{top}, n-3
MV_{bottom}, n-1
MV_{left}, n-5
MV_{right}, n-1

List 2

MV_{zero}, n
$S(MV_{top})$, n-2
$S(MV_{bottom})$, n
$S(MV_{left})$, n-4
$S(MV_{right})$, n

(b) Refined inter MB concealment for MSVC

Fig. 5. Inter MB concealment methods

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 (Varsa et al., 2001). 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. 5(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.

In MSVC, even frames and odd frames are encoded independently, which means the correlation between the reference frame and current frame is reduced. If we only use the frames in the same description as reference frames to conceal the lost MB, it may not perform as well as using reference frames from the other description for concealment. Therefore, we propose to explore the 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. 5(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. 5(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 is corresponding 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

$$S(MV_{candidate}) = \frac{n_c - n_2}{n_c - n_1} MV_{candidate} \tag{3}$$

where n_c is the current frame number. Similarly, motion-compensated concealment is applied
based on the scaled MVs and reference frames in 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.

4. Rate-distortion optimized mode Selection for MSVC

In this section, we propose to estimate the end-to-end distortion for MSVC and incorporate the
estimated distortion with optimal mode selection to improve the robustness of MSVC under
both random and burst losses.

4.1 Related work

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 (Zhu & Kerofsky,
1998; Hinds et al., 1998; Cote & Kossentini, 1999; Cote et al., 2000; Zhang, Regunathan & Rose,
2000; Stockhammer et al., 2002; Eisenberg et al., 2006; Zhang et al., 2007), reference picture
selection (Wiegand et al., 2000; Budagavi & Gibson, 2001), and multiple description video
coding (Wang, Reibman & Lin, 2005).
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 (Zhu & Kerofsky, 1998), or intra-coding blocks
randomly (Cote & Kossentini, 1999).
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 (Hinds et al., 1998; Cote & Kossentini, 1999; Stockhammer et al., 2002;
Cote et al., 2000; Zhang, Regunathan & Rose, 2000; Eisenberg et al., 2006; Zhang et al., 2007).
An early work of RD-based mode selection method is proposed in (Hinds et al., 1998), in
which the distortion is roughly estimated. In Cote & Kossentini (1999), 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 Cote et al. (2000), 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 Hinds et al. (1998), Cote & Kossentini (1999), and Cote et al. (2000) 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 Stockhammer et al. (2002). 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 Zhang, Regunathan & Rose (2000), 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. (2006) estimates the variance of expected distortion by calculating the first four moments of each pixel and incorporates these moments to allocate channel resources. In Zhang et al. (2007), the overall distortion is divided into several separable distortion items to reduce the computing complexity. Heng et al. (2006) 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, Liang et al. (2008) 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 loss has 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 that combines rate-distortion optimized mode selection with MSVC to enhance the error resilience of video (Liao & Gibson, 2009).

4.2 Preliminaries

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 (Wiegand et al., 1996). That is, the coding mode that minimizes the Lagrangian cost in the following equation is chosen to encode the MB,

$$\min_{mode}(J_{MB}) = \min_{mode}(D_{MB} + \lambda_{mode}R_{MB}) \tag{4}$$

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. (5) in H.264,

$$\lambda_{mode} = 0.85 \times 2^{(QP-12)/3} \tag{5}$$

where QP is the quantization parameter of the MB.

f_n^i

$\hat{d}_n^i \leftarrow \hat{f}_n^i$

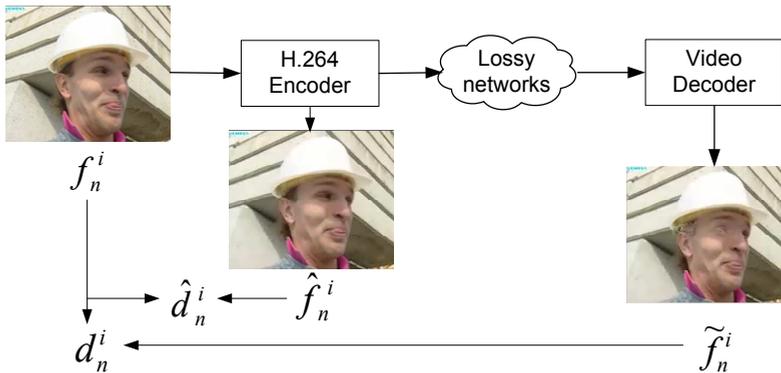$d_n^i \leftarrow \qquad\qquad\qquad\qquad\qquad\qquad \widetilde{f}_n^i$

Fig. 6. 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. 6. 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. 6.

Table 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,

$$D_{MB} = \sum_{i \in MB} d_n^i \qquad\qquad (6)$$

In the next section, we derive an approach to estimate the distortion d_n^i for MSVC by considering the quantization, packet losses, error propagation, and multiple state recovery.

4.3 Estimation of end-to-end distortion for MSVC

From Section 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

	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)

Table 1. Notation

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 on a MB basis are applied (Liao & Gibson, 2008). 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.

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

$$d_n^i = E[(f_n^i - \tilde{f}_n^i)^2] = (f_n^i)^2 - 2f_n^i E[\tilde{f}_n^i] + E[(\tilde{f}_n^i)^2] \tag{7}$$

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 for MSVC.

4.3.1 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 $m = \underset{x \in \{1,2\}}{\arg\min}(\tilde{f}_{n-x}^{j_x} - \hat{f}_n^i)^2$.

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 $k = \underset{x \in \{1,2\}}{\arg\min}(\tilde{f}_{n-x}^i - \hat{f}_n^i)^2$.

Based on the above cases, we can calculate the first and second moments of $\tilde{f}_n^i$ in an intra MB by Eqs. (8) and (9),

$$E[\tilde{f}_n^i] = (1 - p)(\hat{f}_n^i) + p(1 - p)E[\tilde{f}_{n-m}^{j_m}] + p^2 E[\tilde{f}_{n-k}^i] \tag{8}$$

$$E[(\tilde{f}_n^i)^2] = (1 - p)(\hat{f}_n^i)^2 + p(1 - p)E[(\tilde{f}_{n-m}^{j_m})^2] + p^2 E[(\tilde{f}_{n-k}^i)^2] \tag{9}$$

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

4.3.2 Pixel in an inter-coded MB

For MSVC, the odd frame is predicted from previous odd frames and the even frame is predicted from 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 $n-2$. 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 4.3.1,

$$E[\tilde{f}_n^i] = (1-p)(\hat{r}_n^i + E[\tilde{f}_{n-2}^j]) + p(1-p)E[\tilde{f}_{n-m}^{j_m}] + p^2 E[\tilde{f}_{n-k}^i] \tag{10}$$

$$E[(\tilde{f}_n^i)^2] = (1-p)E[(\hat{r}_n^i + \tilde{f}_{n-2}^j)^2] + p(1-p)E[(\tilde{f}_{n-m}^{j_m})^2] + p^2 E[(\tilde{f}_{n-k}^i)^2] \tag{11}$$

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

5. Experimental results

5.1 Simulation setup

In this section, we discuss the model we used to simulate the packet losses in wireless networks, the performance metrics to evaluate performance of different methods, and the parameters and methods to encode the video for comparison.

5.1.1 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. 7, 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 identically 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:
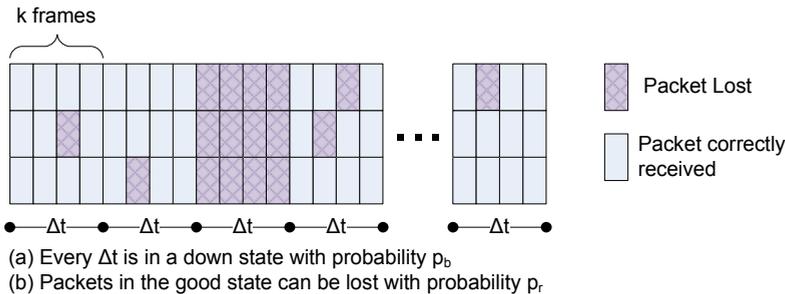

(a) Every Δt is in a down state with probability p_b
(b) Packets in the good state can be lost with probability p_r

Fig. 7. Packet Loss Model

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

$$p = p_b + (1 - p_b)p_r = p_b + p_r - p_b p_r \tag{12}$$

5.1.2 Performance metrics

To analyze the performance of the decoded video sequences, we use the average peak signal-to-noise ratio (PSNR) of all frames over all realizations to evaluate the objective video quality, because it is the most widely used objective video quality metric. PSNR represents the mean squared errors (MSE) of the distorted videos and is defined by

$$PSNR = 10 \cdot \log_{10} \frac{255^2}{MSE} \tag{13}$$

where MSE is the mean square error between the original pixel and the distorted pixel. However, due to the non-linear behavior of human visual system, video sequences with close average PSNR may reveal different perceptual video quality for human viewers. Therefore, we also introduce $PSNR_{r,f}$ proposed by Hu et al. (2008) to evaluate the perceptual video quality for multiple channel uses. $PSNR_{r,f}$ 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 $PSNR_{r,f}$ can be written as

$$PSNR_{r,f} = \arg_x P_{real}(P_{frame}(PSNR > x) \geq f\%) \geq r\%) \tag{14}$$

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, $PSNR_{r=80\%, f=90\%} = 35$ dB means that there are 80% of the realizations having 90% of frames with PSNR higher than 35 dB. We use $PSNR_{r,f}$ as a multiuser perceptual video quality indicator because of two reasons. First, $PSNR_{r,f}$ 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 (Wang et al., 2003): (1) The bad-quality frames in a video dominates 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, $PSNR_{r,f}$ captures the performance loss due to damaged frames in a video sequence ($f\%$). Second, due to the time-variant network conditions, multiple users or a user in multiple uses may have different experience. $PSNR_{r,f}$ can capture the performance experienced by a user in multiple uses (r%), or alternatively, it indicates the percentage of video users that experience a specific video quality.

5.1.3 Compared schemes and simulation settings

We implement our proposed methods by modifying H.264 reference software JM13.2. The temporal copy method is used to estimate the end-to-end distortion for mode selection. To evaluate the performance of the refined error concealment method for MSVC, we compare the following three coding schemes:

1. **SDC**: The video sequence is coded into a single description and transmitted over one path over the network.

2. **MSVC**: The video sequence is coded into two descriptions using MSVC and transmitted over two independent paths over the network.

3. **MSVC_REC**: The video sequence is coded and transmitted the same as MSVC, while the refined error concealment method introduced in Section 3 is applied to decode the corrupted video.

Furthermore, we investigate the performance of MSVC with the optimal mode selection and refined error concealment methods by comparing the following three coding schemes:

1. **SDC_ROPE**: The video sequence is coded into a single description with ROPE proposed in Zhang, Regunathan & Rose (2000) and transmitted over one path over the network.

2. **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 is applied to decode the corrupted video.

3. **MSVC_OMS**: The video sequence is coded into two descriptions using MSVC with the optimal mode selection introduced in Section 4. Then the encoded bitstream is transmitted over two paths and decoded using the refined error concealment method for MSVC.

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

$$l = \frac{r}{8f \cdot n} (bytes/packet) \tag{15}$$

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 (Wenger, 2003). The packet loss model in Section 5.1.1 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.

5.2 Performance evaluation
5.2.1 Refined error concealment for MSVC

In Fig. 8, we show the PSNR performance under different bitrates for Foreman sequence. Figure 8(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. Therefore, MSVC outperforms SDC at a higher packet loss rate as shown in Fig. 8(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. 8(a)) and in the range of 1.4-2.8 dB at a 8% packet loss rate (Fig. 8(b)). 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 Fig. 9(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 SDC is more vulnerable to burst losses. That is

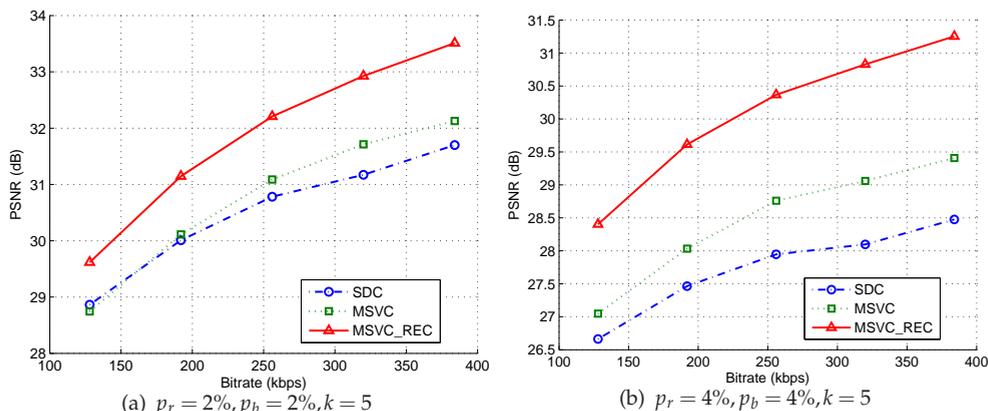(a) $p_r = 2\%, p_b = 2\%, k = 5$ (b) $p_r = 4\%, p_b = 4\%, k = 5$

Fig. 8. Average PSNR vs Bitrate for SDC, MSVC, and MSVC_REC, Foreman sequence at 30 fps

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 9(a) shows that MSVC and MSVC_REC are more effective to combat burst losses than SDC. Meanwhile, MSVC_REC has higher PSNR than MSVC of about 0.6 dB under various burst loss rates.

Figure 9(b) investigates the PSNR performance of SDC, MSVC, and MSVC_REC under different random loss rates with a fixed burst loss rate at 1%. 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.

In Table 2, we compare SDC, MSVC, and MSVC_REC for different video sequences. The results demonstrate that MSVC_REC can provide better reconstructed video quality for video communications over lossy networks.
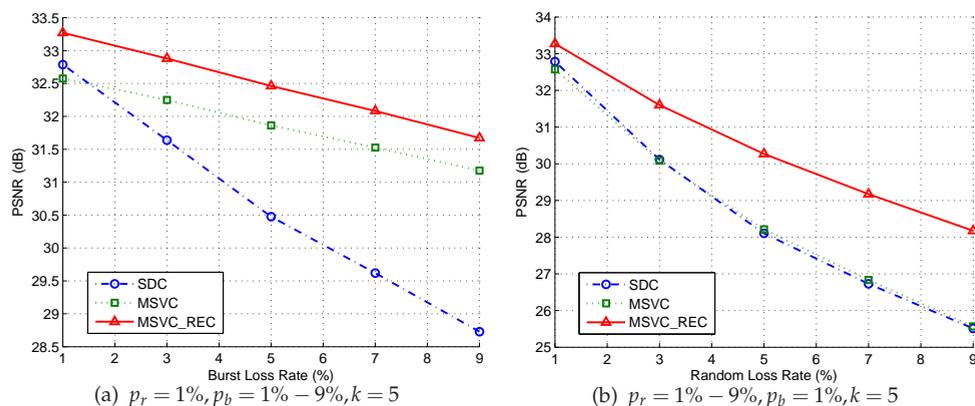

(a) $p_r = 1\%, p_b = 1\% - 9\%, k = 5$ (b) $p_r = 1\% - 9\%, p_b = 1\%, k = 5$

Fig. 9. Average PSNR vs Packet Loss Rate for SDC, MSVC, and MSVC_REC, Foreman sequence at 30 fps, 256 kbps

(a) $p_r = 2\%, p_b = 2\%, k = 5$ (b) $p_r = 4\%, p_b = 4\%, k = 5$
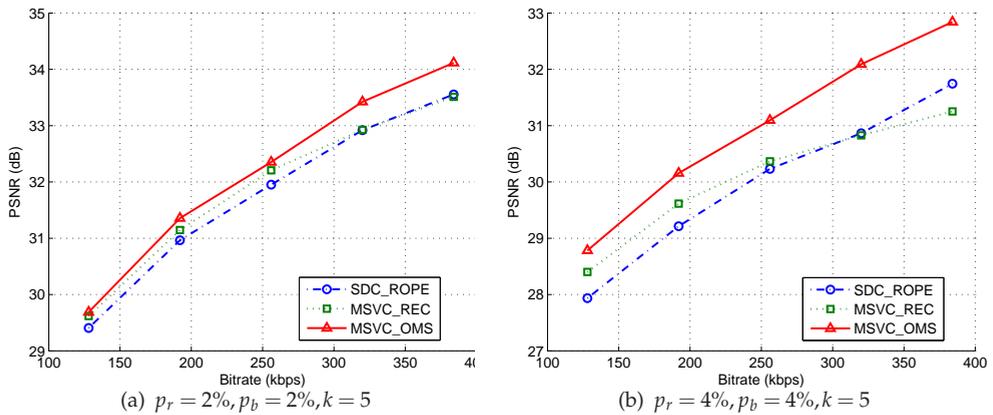
Fig. 10. Average PSNR vs Bitrate for SDC_ROPE, MSVC_REC, and MSVC_OMS, Foreman sequence at 30 fps

5.2.2 Rate-distortion optimized mode selection for MSVC

In this section, we evaluate the performance of our optimal mode selection method for MSVC. First, we show the average PSNRs of SDC_ROPE, MSVC_REC, and MSVC_OMS under different bitrates for two sets of network conditions in Fig. 10.

We see that MSVC_OMS achieves more gain than MSVC_REC at a higher packet loss rate as shown in Fig. 10(b), since MSVC_OMS adaptively selects the optimal coding mode based on the packet loss rate in the network. Figure 10 also shows that MSVC_OMS has the best PSNR performance across different bitrates.

We have shown the objective performance of three methods in Fig. 10. Now we look into a case to investigate the perceptual video quality of the three methods. Figure 11 shows $PSNR_{r,f}$ of SDC_ROPE, and MSVC_REC, and MSVC_OMS for Foreman sequence under network condition ($p_r = 4\%, p_b = 4\%, k = 5$). The average PSNR of these three methods are 30.23 dB, 30.37 dB, and 31.09 dB respectively. Figure 11(a) presents the $PSNR_{r,f}$ values with a fixed $r = 85\%$. Compared to SDC_ROPE and MSVC_REC, MSVC_OMS has the fewest number of low-quality frames in 85% of the realizations. For example, as shown in Fig. 11(a), about 25% of 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.

Sequence	$p_b = 2\%, p_r = 2\%, k = 5$			$p_b = 4\%, p_r = 4\%, k = 5$		
	SDC	MSVC	MSVC_REC	SDC	MSVC	MSVC_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

Table 2. Average PSNR (dB) for different video sequences, 30 fps, 256 kbps
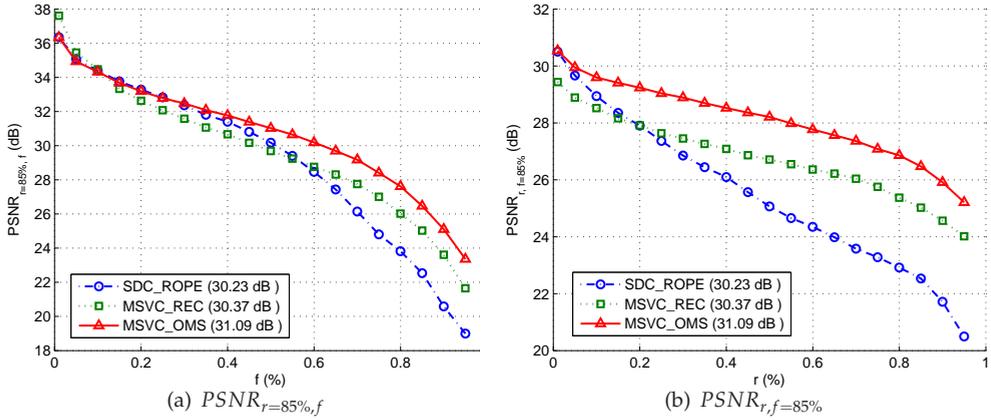
Fig. 11. $PSNR_{r,f}$ for Foreman sequence at 30 fps, 256 kbps, $p_r = 4\%$, $p_b = 4\%$, $k = 5$

Figure 11(b) plots $PSNR_{r,f}$ of SDC_ROPE, MSVC_REC, and MSVC_OMS with fixed $f = 85\%$. In the figure, we see that MSVC_OMS has the highest $PSNR_{r,f=85\%}$ under most values of r. This means that MSVC_OMS can guarantee a higher PSNR than SDC_ROPE 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, and MSVC_OMS are 22.53 dB, 25.02 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.

Table 3 presents the average PSNR and $PSNR_{r=85\%,f=85\%}$ results for SDC_ROPE, MSVC_REC, and MSVC_OMS for different video sequences. The results show that MSVC_OMS not only provides best objective video quality among the three methods but also has the best perceptual performance for multiple users.

6. Conclusion

We explore approaches to combine source coding diversity with path diversity to support video communications over wireless ad-hoc networks. There are several benefits of transmitting multiple independent source descriptions over different paths:

1. Traffic dispersion and load balancing: Sending a video bitstream across multiple paths reduces the per-path bandwidth, thus relieves congestion at hotspots and enhances network utilization.
2. Reduced burst losses: Distributing video packets through multiple paths increases the time interval to send video packets on each path, therefore, for a given duration of congestion, fewer packets are lost.
3. Improved error resilience: As long as the descriptions do not have simultaneous losses, the system can reconstruct the video with acceptable video quality.

In this article, we focus on a source coding diversity method called MSVC, because this method is easy to implement and compatible with different video standards. We propose methods at both the encoder and decoder sides to improve the error robustness of MSVC. The

Sequence	$p_b = 2\%, p_r = 2\%, k = 5$					
	SDC_ROPE	MSVC_REC	MSVC_OMS	SDC_ROPE	MSVC_REC	MSVC_OMS
	Average PSNR (dB)			$PSNR_{r=85\%, f=85\%}$		
Carphone	33.23	33.21	33.88	23.97	27.86	28.50
Claire	41.37	42.31	42.66	27.22	37.31	38.84
Foreman	31.95	32.21	32.35	24.29	27.99	28.64
Hall-monitor	38.40	38.46	39.09	25.76	32.58	33.77
Mother-daughter	38.90	39.20	39.45	30.69	35.91	36.82
News	37.84	37.08	37.31	25.21	30.58	31.39
	$p_b = 4\%, p_r = 4\%, k = 5$					
	Average PSNR (dB)			$PSNR_{r=85\%, f=85\%}$		
Carphone	31.25	31.02	32.38	21.63	25.17	26.31
Claire	38.68	40.42	41.17	24.40	32.88	35.28
Foreman	30.23	30.37	31.09	22.53	25.02	26.47
Hall-monitor	35.66	36.52	37.72	22.17	29.21	29.60
Mother-daughter	37.25	37.89	38.38	26.82	33.18	34.40
News	35.26	35.20	35.80	22.67	26.66	26.98

Table 3. Average PSNR and $PSNR_{r=85\%, f=85\%}$ for different video sequences, 30 fps, 256 kbps

encoder estimates the end-to-end distortion for MSVC by considering the network conditions, error propagation, and error concealment, then applies RD-based mode selection to select the optimal coding mode. This helps to alleviate error propagation due to packet loss in MSVC. The decoder applies refined error concealment methods for MSVC by using the information available across two descriptions to recover lost MBs.

We study the performance of our proposed methods over a wireless network with both random and burst losses. The results show that our proposed methods have PSNR gain between 0.3 dB and 2.9 dB under various bitrates and network conditions for different video sequences. We also use a multiuser perceptual video quality indicator $PSNR_{r,f}$ to evaluate the perceptual performance for multiple users. The results demonstrate that our proposed methods provide a better video experience for multiple users in the network.

7. References

Aign, S. & Fazel, K. (1995). Temporal and spatial error concealment techniques for hierarchical MPEG-2 video codec, *IEEE International Conference on Communications*, Vol. 3, pp. 1778–1783.

Apostolopoulos, J. G. (2001). Reliable video communication over lossy packet networks using multiple state encoding and path diversity, *SPIE Proceedings on Visual Communications and Image Processing*, Vol. 4310, pp. 392–409.

Apostolopoulos, J. G. & Trott, M. D. (2004). Path diversity for enhanced media streaming, *IEEE Communications Magazine* **42**(8): 80–87.

Atzori, L., De Natale, F. & Perra, C. (2001). A spatio-temporal concealment technique using boundary matching algorithm and mesh-based warping (BMA-MBW), *IEEE Transactions on Multimedia* **3**(3): 326–338.

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

Budagavi, M. & Gibson, J. D. (2001). Multiframe video coding for improved performance over wireless channels, *IEEE Transactions on Image Processing* **10**(2): 252 –265.

Chen, Y., Hu, Y., Au, O. C., Li, H. & Chen, C. W. (2008). Video error concealment using spatio-temporal boundary matching and partial differential equation, *IEEE Transactions on Multimedia* **10**(1): 2–15.

Cote, G. & Kossentini, F. (1999). Optimal intra coding of blocks for robust video communication over the internet, *Signal Processing: Image Communication*, Vol. 15, pp. 25–34.

Cote, G., Shirani, S. & Kossentini, F. (2000). Optimal mode selection and synchronization for robust video communications over error-prone networks, *IEEE Journal on Selected Areas in Communications* **18**(6): 952–965.

Dumitrescu, S. & Wu, X. (2009). On properties of locally optimal multiple description scalar quantizers with convex cells, *IEEE Transactions on Information Theory* **55**(12): 5591–5606.

Eisenberg, Y., Zhai, F., Pappas, T. N., Berry, R. & Katsaggelos, A. K. (2006). VAPOR: variance-aware per-pixel optimal resource allocation, *Image Processing, IEEE Transactions on* **15**(2): 289–299.

Franchi, N., Fumagalli, M., Lancini, R. & Tubaro, S. (2005). Multiple description video coding for scalable and robust transmission over IP, *IEEE Transactions on Circuits and Systems for Video Technology* **15**(3): 321–334.

Gogate, N., Chung, D.-M., Panwar, S. S. & Wang, Y. (2002). 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* **12**(9): 777–792.

Goyal, V. K. (2001). Multiple description coding: compression meets the network, *IEEE Signal Processing Magazine* **18**(5): 74–93.

Haskell, P. & Messerschmitt, D. (1992). Resynchronization of motion compensated video affected by ATM cell loss, *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 3, pp. 545–548.

Hemami, S. S. & Meng, T. H.-Y. (1995). Transform coded image reconstruction exploiting interblock correlation, *IEEE Transactions on Image Processing* **4**(7): 1023–1027.

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

Hinds, R. O., Pappas, T. N. & Lim, J. S. (1998). Joint block-based video source/channel coding for packet-switched networks, *Visual Communications and Image Processing*, Vol. 3309, SPIE, pp. 124–133.

Hu, J., Choudhury, S. & Gibson, J. D. (2008). Video capacity of WLANs with a multiuser perceptual quality constraint, *IEEE Transactions on Multimedia* **10**(8): 1465–1478.

Hyun, S. H., Kim, S. S., Kim, B. C., Eom, I. K. & Kim, Y. S. (2008). Efficient directional interpolation for block recovery using difference values of border pixels, *Congress on Image and Signal Processing*, Vol. 3, pp. 565–568.

Kung, W.-Y., Kim, C.-S. & Kuo, C.-C. (2006). Spatial and temporal error concealment techniques for video transmission over noisy channels, *IEEE Transactions on Circuits*

and Systems for Video Technology **16**(7): 789–803.

Kwok, W. & Sun, H. (1993). Multi-directional interpolation for spatial error concealment, *IEEE Transactions on Consumer Electronics* **39**(3): 455–460.

Lam, W. M., Reibman, A. R. & Liu, B. (1993). Recovery of lost or erroneously received motion vectors, *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 5, pp. 417–420.

Lee, H. Y., Eom, I. K. & Kim, Y. S. (2004). Error concealment using directional coefficient mask and difference of DC, *30th Annual Conference of IEEE Industrial Electronics Society*, Vol. 3, pp. 2086–2091.

Lee, X., Zhang, Y.-Q. & Leon-Garcia, A. (1995). Information loss recovery for block-based image coding techniques-a fuzzy logic approach, *IEEE Transactions on Image Processing* **4**(3): 259–273.

Lee, Y.-C. & Altunbasak, Y. (2002). A collaborative multiple description transform coding and statistical error concealment method for error resilient video streaming over noisy channels, *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 2, pp. 2077–2080.

Liang, Y. J., Apostolopoulos, J. G. & Girod, B. (2008). 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* **18**(7): 861–874.

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

Liao, Y. & Gibson, J. D. (2009). Rate-distortion based mode selection for video coding over wireless networks with burst losses, *17th International Packet Video Workshop*, pp. 1–10.

Lim, K. P., Sullivan, G. & Wiegand, T. (2005). 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* **6**.

Lu, Y., Zhou, R., Cui, H. & Tang, K. (2005). Bi-directional entire frame recovery in MDC video streaming, *IEEE International Symposium on Communications and Information Technology*, Vol. 2, pp. 1058–1061.

Luo, H.-y., Gan, Z.-l. & Zhu, X.-c. (2009). Content-adaptive interpolation for spatial error concealment, *Second International Conference on Information and Computing Science*, Vol. 1, pp. 265–268.

Ma, M., Au, O. C., Guo, L., Chan, S.-H. G. & Wong, P. H. W. (2008). Error concealment for frame losses in MDC, *IEEE Transactions on Multimedia* **10**(8): 1638–1647.

Mao, S., Lin, S., Panwar, S. S., Wang, Y. & Celebi, E. (2003). Video transport over ad hoc networks: multistream coding with multipath transport, *IEEE Journal on Selected Areas in Communications* **21**(10): 1721–1737.

Park, J. W., Kim, J. W. & Lee, S. U. (1997). 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* **7**(6): 845–854.

Reibman, A., Jafarkhani, H., Wang, Y. & Orchard, M. (2001). Multiple description video using rate-distortion splitting, *Proceedings of International Conference on Image Processing*, Vol. 1, pp. 978–981.

Salama, P., Shroff, N. B. & Delp, E. J. (2000). Error concealment in MPEG video streams over ATM networks, *IEEE Journal on Selected Areas in Communications* **18**(6): 1129–1144.

Shirani, S., Kossentini, F. & Ward, R. (2000). A concealment method for video communications

in an error-prone environment, *IEEE Journal on Selected Areas in Communications*
18(6): 1122–1128.

Stockhammer, T., Kontopodis, D. & Wieg, T. (2002). Rate-distortion optimization for
JVT/H.26L video coding in packet loss environment, *Proceedings of International
Packet Video Workshop*.

Sun, H. & Kwok, W. (1995). Concealment of damaged block transform coded images using
projections onto convex sets, *IEEE Transactions on Image Processing* **4**(4): 470–477.

Tsekeridou, S. & Pitas, I. (2000). MPEG-2 error concealment based on block-matching
principles, *IEEE Transactions on Circuits and Systems for Video Technology* **10**(4): 646
–658.

Vaishampayan, V. (1993). Design of multiple description scalar quantizers, *IEEE Transactions
on Information Theory* **39**(3): 821–834.

Varsa, V., Hannuksela, M. & Wang, Y. (2001). Non-normative error concealment algorithms,
ITU-T, VCEG-N62 .

Wang, D., Canagarajah, N., Agrafiotis, D. & Bull, D. (2005). Error concealment for slice
group based multiple description video coding, *IEEE International Conference on Image
Processing*, Vol. 1, pp. I–769–772.

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

Wang, Y., Orchard, M. T., Vaishampayan, V. & Reibman, A. R. (2001). Multiple description
coding using pairwise correlating transforms, *IEEE Transactions on Image Processing*
10(3): 351–366.

Wang, Y., Reibman, A. R. & Lin, S. (2005). Multiple description coding for video delivery,
Proceedings of the IEEE **93**(1): 57–70.

Wang, Y. & Zhu, Q.-F. (1998). Error control and concealment for video communication: a
review, *Proceedings of the IEEE* **86**(5): 974–997.

Wang, Y., Zhu, Q.-F. & Shaw, L. (1993). Maximally smooth image recovery in transform
coding, *IEEE Transactions on Communications* **41**(10): 1544–1551.

Wang, Z., Sheikh, H. R. & Bovik, A. C. (2003). Objective video quality assessment, *The
Handbook of Video Databases: Design and Applications*, CRC Press, pp. 1041–1078.

Wenger, S. (2003). H.264/AVC over IP, *IEEE Transactions on Circuits and Systems for Video
Technology* **13**(7): 645–656.

Wiegand, T., Farber, N., Stuhlmuller, K. & Girod, B. (2000). Error-resilient video transmission
using long-term memory motion-compensated prediction, *IEEE Journal on Selected
Areas in Communications* **18**(6): 1050–1062.

Wiegand, T., Lightstone, M., Mukherjee, D., Campbell, T. G. & Mitra, S. K. (1996).
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* **6**(2): 182–190.

Wiegand, T., Sullivan, G. J., Bjontegaard, G. & Luthra, A. (2003). Overview of the H.264/AVC
video coding standard, *IEEE Transactions on Circuits and Systems for Video Technology*
13(7): 560–576.

Wu, J., Liu, X. & Yoo, K.-Y. (2008). A temporal error concealment method for H.264/AVC using
motion vector recovery, *IEEE Transactions on Consumer Electronics* **54**(4): 1880–1885.

Zeng, W. & Liu, B. (1999). Geometric-structure-based error concealment with novel
applications in block-based low-bit-rate coding, *IEEE Transactions on Circuits and*

Systems for Video Technology **9**(4): 648–665.

Zhang, J., Arnold, J. F. & Frater, M. R. (2000).   A cell-loss concealment technique for MPEG-2 coded video, *IEEE Transactions on Circuits and Systems for Video Technology* **10**(4): 659–665.

Zhang, R., Regunathan, S. L. & Rose, K. (2000). Video coding with optimal inter/intra-mode switching for packet loss resilience, *IEEE Journal on Selected Areas in Communications* **18**(6): 966–976.

Zhang, R., Zhou, Y. & Huang, X. (2004). Content-adaptive spatial error concealment for video communication, *IEEE Transactions on Consumer Electronics* **50**(1): 335–341.

Zhang, Y., Gao, W., Lu, Y., Huang, Q. & Zhao, D. (2007). Joint source-channel rate-distortion optimization for H.264 video coding over error-prone networks, *IEEE Transactions on Multimedia* **9**(3): 445–454.

Zhou, Z.-H., Xie, S.-L. & Xu, Z.-L. (2005). Efficient adaptive MRF-MAP error concealment of video sequences, *Proceedings of International Conference on Machine Learning and Cybernetics*, Vol. 9, pp. 5507–5511.

Zhu, Q.-F. & Kerofsky, L. (1998).   Joint source coding, transport processing, and error concealment for H.323-based packet video, *Visual Communications and Image Processing*, Vol. 3653, SPIE, pp. 52–62.