Error Resilient Coding and Error Concealment in Scalable Video Coding

Yi Guo, Ying Chen, *Member, IEEE*, Ye-Kui Wang, *Member, IEEE*, Houqiang Li, Miska M. Hannuksela, *Member, IEEE*, and Moncef Gabbouj, *Senior Member, IEEE*

***Abstract*— Scalable video coding (SVC), which is the scalable extension of the H.264/AVC standard, was developed by the Joint Video Team (JVT) of ISO/IEC MPEG (Moving Picture Experts Group) and ITU-T VCEG (Video Coding Experts Group). SVC is designed to provide adaptation capability for heterogeneous network structures and different receiving devices with the help of temporal, spatial, and quality scalabilities. It is challenging to achieve graceful quality degradation in an error-prone environment, since channel errors can drastically deteriorate the quality of the video. Error resilient coding and error concealment techniques have been introduced into SVC to reduce the quality degradation impact of transmission errors. Some of the techniques are inherited from or applicable also to H.264/AVC, while some of them take advantage of the SVC coding structure and coding tools. In this paper, the error resilient coding and error concealment tools in SVC are first reviewed. Then, several important tools such as loss-aware rate-distortion optimized macroblock mode decision algorithm and error concealment methods in SVC are discussed and experimental results are provided to show the benefits from them. The results demonstrate that PSNR gains can be achieved for the conventional inter prediction (IPPP) coding structure or the hierarchical bi-predictive (B) picture coding structure with large group of pictures size, for all the tested sequences and under various combinations of packet loss rates, compared with the basic Joint Scalable Video Model (JSVM) design applying no error resilient tools at the encoder and only picture copy error concealment method at the decoder.**

***Index Terms*— Error concealment, error resilient coding, H.264/AVC, SVC.**

I. INTRODUCTION

SCALABLE VIDEO coding, also referred to as layered video coding, has been designed to facilitate video services using a single bit stream, from which appropriate sub-bit stream can be extracted to meet different preferences and requirements for a possibly large number of end users, over heterogeneous network structures with a wide range of quality of service (QoS). In scalable video coding (SVC), a video is coded into more than one layer: the base layer and enhancement layers, the latter of which usually can improve user experience with respect to picture rate, spatial resolution, and/or video quality. These enhancements are referred to as temporal, spatial, and SNR scalabilities, respectively, and can be used in a combined manner.

Manuscript received March 29, 2008; revised June 21, 2008. First version published March 16, 2009; current version published June 19, 2009. This paper is partially supported by National Natural Science Foundation of China (NSFC) General Program (Contract No. 60572067&60672161), and NSFC Key Program (Contract No. 60736043). It is also supported in part by Nokia and the Academy of Finland, Finnish Center of Excellence Program 2006-2011 under Project 213462. This paper was recommended by Associate Editor H. Sun.

Y. Guo and H. Li are with the Department of Electronic Engineering and Information Science at the University of Science and Technology of China, Hefei (e-mail: guoyi@mail.ustc.edu.cn; lihq@ustc.edu.cn).

Y. Chen and M. Gabbouj are with the Department of Signal Processing at Tampere University of Technology, Tampere, Finland (e-mail: ying.chen@tut.fi; moncef.gabbouj@tut.fi).

Y.-K. Wang and M. M. Hannuksela are with the Nokia Research Center, Tampere, Finland (e-mail: ye-kui.wang@nokia.com; miska.hannuksela@nokia.com).

Digital Object Identifier 10.1109/TCSVT.2009.2017311

A. Scalable Video Coding Over Heterogeneous Networks

Typical application scenarios for SVC are shown in Fig. 1. Note that, in this figure, only spatial and temporal scalabilities are shown. However, the scenarios for spatial scalability are also valid for SNR scalability. In practice, those scenarios may exist in different systems with different contents, network structures, and receiving devices.

Due to various levels of decoding capability, videos with different spatial resolutions, e.g., for a standard definition TV (SDTV) set and a high definition TV (HDTV) set, can be decoded as shown in scenario (a), or videos with different picture rates, e.g., for a mobile device and a laptop, can be decoded as shown in scenario (b).

The clients can be the same but within different subnetworks or with different connections, e.g. in scenario (c). The clients are connected with cable, local area network (LAN), digital subscriber line (DSL), and wireless LAN (WLAN). Clients can also be located in the same network but with different QoS, e.g., the different congestion control methods applied by the intermediate nodes. Therefore, the expected bandwidth for each client may be different, which will lead to various received videos combined with different picture rates, spatial resolutions, and/or quality levels.

Even for one client, owing to bandwidth fluctuation, the received video may change at any moment in picture rate, spatial resolution, and quality level.

B. Error Robust Requirement and Error Control

The number of packet-based video transmission channels, such as the Internet and packet-oriented wireless networks, has been increasing rapidly. One inherent problem of video transmitted in packet-oriented transport protocol is channel errors, as client 4 in scenario (c) of Fig. 1. Packet loss may be caused if a packet fails to reach the destination in a specific time. Another source of packet loss is bit errors caused by physical interference in any link of the transmission

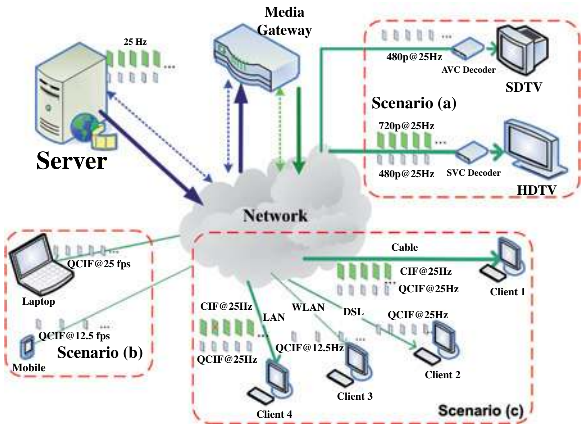


Fig. 1. Scalable video coding application scenarios.

path. Many video communication systems apply the user datagram protocol (UDP) [1]. Any bit error occurring in a UDP packet will result in the loss of the packet, as UDP discards packets with bit errors. Packet loss can damage one whole picture or an area of it. Unfortunately, because of the predictive coding techniques, a transmission error (after error concealment) will propagate both temporally and spatially, and sometimes can bring substantial deterioration to the subjective and objective quality of the reproduced video sequence until an instantaneous decoding refresh (IDR) picture. However, if the bit stream is protected by error control methods [2], the system may still maintain graceful degradation.

Various error control methods have been proposed. In [3], error control methods are classified into four types as follows: transport-level error control; source-level error resilient coding; interactive error control; and error concealment.

This paper will mainly focus on source-level error resilient coding and error concealment. Error resilient coding injects such redundancy into the bit stream, which helps receivers in recovery or concealment from potential channel errors. The objective of error resilient coding is to design a scheme that can achieve the minimum end-to-end distortion under a certain rate. The redundancy may be used to detect data losses, stop error propagation, and/or guide error concealment. Error concealment provides an estimation of lost picture areas based on the correctly decoded samples as well as any other helpful information. Error concealment is done only by the decoder, unlike other methods that require encoder actions.

C. Outline and Contribution of This Paper

In this paper, error resilient coding and error concealment techniques used in single-layer coding are reviewed first. Some of these techniques are included in or can be applied to SVC [4], which is the scalable extension of H.264/AVC [5]. Several new error resilient techniques in SVC, including some normative tools as well as the non-normative loss-aware rate-distortion optimized mode decision (LA-RDO) algorithm, are then discussed. Furthermore, error concealment algorithms, which are designed according to new characteristics of SVC, e.g., inter-layer texture, motion and residual prediction, are discussed. It is shown that techniques based on the inter-layer correlation can outperform the techniques inherited from single-layer coding, only based on the spatial/temporal correlations.

The rest of this paper is organized as follows. First, an overview of SVC is given in Section II in order to help understand the discussion of the error resilient coding and error concealment tools. In Section III, techniques for single-layer coding, especially for H.264/AVC, are introduced. The error resilient coding and error concealment tools, most of which were proposed by the authors of this paper, are discussed in Section IV. Simulation results are provided in Section V to show the benefits of the proposed algorithms. Finally, Section VI concludes the paper.

II. Overview of the Scalable Extension of H.264/AVC

This section reviews SVC (the scalable extension of H.264/AVC), which is important to understand the terminologies required for SVC error resilient coding and error concealment. SVC has been included in MPEG-2 video (also known as ITU-T H.262) [6], H.263 [7], MPEG-4 visual [8], and SVC, which all provide temporal, spatial, and SNR scalabilities.

A. Novel Features of SVC

Some functionalities of SVC are inherited from H.264/AVC. Compared to previous scalable standards, the most essential advantages, namely hierarchical temporal scalability, inter-layer prediction, single-loop decoding, and flexible transport interface, are reviewed below.

According to the SVC specification, the pictures with the lowest spatial and quality layer are compatible with H.264/AVC, and their pictures of the lowest temporal level form the temporal base layer, which can be enhanced with pictures of higher temporal levels. In addition to the H.264/AVC-compatible layer, several spatial and/or SNR enhancement layers can be added to provide spatial and/or quality scalabilities. SNR scalability is also referred to as quality scalability. Each spatial or SNR enhancement layer itself may be temporally scalable, with the same temporal scalability structure as the H.264/AVC-compatible layer. For one spatial or SNR enhancement layer, the lower layer it depends on is also referred to as the base layer of that specific spatial or SNR enhancement layer. In this paper, unless otherwise stated, the term "base layer" refers to a certain spatial or SNR layer, the information (texture, residue, and motion) of which may be used as inter-layer prediction by a higher spatial or SNR layer, and the term "enhancement layer" refers to the specific higher spatial or SNR layer.

1) Hierarchical Temporal Scalability: H.264/AVC provides a flexible hierarchical B picture coding structure, which enables it to realize advanced temporal scalability [9]. With this feature inherited from H.264/AVC, SVC supports temporal scalability for layers with different resolutions [10]. In SVC, a group of pictures (GOP) consists of a so-called key picture and all pictures that are located in output/display order between this key picture and the previous key picture. A key picture is coded in regular or irregular intervals, which is either intra-coded or inter-coded using the previous key picture as reference for motion-compensated prediction. The non-key pictures are hierarchically predicted from the pictures with lower temporal levels. The temporal level of a picture is indicated by the syntax element temporal_id in the network abstraction layer (NAL) unit header SVC extension [4].

2) Inter-layer Prediction: SVC introduces inter-layer prediction for spatial and SNR scalabilities based on texture, residue, and motion. The spatial scalability in SVC has been generalized into any resolution ratio between two layers [10]. The SNR scalability can be realized by coarse granularity scalability (CGS) or medium granularity scalability (MGS) [10]. In SVC, two spatial or CGS layers belong to different dependency layers (indicated by dependency_id in NAL unit header [4]), while two MGS layers can be in the same dependency layer. One dependency layer includes quality layers with quality_id [4] from zero to higher values, which correspond to quality enhancement layers. In SVC, inter-layer prediction methods are utilized to reduce the inter-layer redundancy. They are briefly introduced in the following paragraphs.

1) Inter-layer texture prediction: The coding mode using inter-layer texture prediction is called "IntraBL" mode in SVC. To enable single-loop decoding [11], only the macroblocks (MBs) whose co-located MBs in the base layer are constrainedly intra-coded can use this mode. A constrainedly intra-coded macroblock (MB) is intra-coded without referring to any samples from the neighboring MBs that are inter-coded.
2) Inter-layer residual prediction: If an MB is indicated to use residual prediction, the co-located MB in the base layer for inter-layer prediction must be an inter MB and its residue may be upsampled according to the resolution ratio. The difference between the residue of the enhancement layer and that of the base layer is coded.
3) Inter-layer motion prediction: The co-located base layer motion vectors may be scaled to generate predictors for the motion vectors of MB or MB partition in the enhancement layer. In addition, there is one MB type named base mode, which sends one flag for each MB. If this flag is true and the corresponding base layer MB is not intra, then motion vectors, partitioning modes and reference indices are all derived from base layer.

3) Single-loop Decoding: The single-loop decoding scheme in SVC is revolutionary compared to earlier scalable coding techniques. In the single-loop decoding scheme, only the target layer needs to be motion-compensated and fully decoded [11]. Therefore, compared to the conventional multiple-loop decoding scheme, where motion compensation and full decoding are typically required for every spatial or SNR-scalable layer, decoding complexity as well as the decoded picture buffer (DPB) size can be greatly reduced.

4) Flexible Transport Interface: SVC provides flexible systems and transport interface designs that enable seamless integration of the codec to scalable multimedia application systems. Other than compression and scalability provisioning, systems and transport interface focuses on codec functionalities, such as, for video codec in general, interoperability and conformance, extensibility, random access, timing, buffer management, as well as error resilience, and for scalable coding in particular, backward compatibility, scalability information provisioning, and scalability adaptation. These mechanisms are augmented by the SVC file format extension to the International Standardization Organization (ISO) Base Media File Format [12] and Real-time Transport Protocol (RTP) payload formats [13]. Discussions of these SVC systems and transport interface designs can be found in [12], [13], and [14]. The error resilient coding and error concealment tools that are applicable to SVC are discussed in the following sections of this paper.

III. Overview of Error Resilient Coding and Error Concealment Tools for H.264/AVC

Earlier video coding standards (H.261/3, MPEG-1/2/4) support the following standard error resilient coding tools: 1) intra MB/picture refresh [15]; 2) slice coding [15]; 3) reference picture identification (see below); 4) reference picture selection (RPS) [15]; 5) data partitioning [15]; 6) header extension code and header repetition [15]; 7) spare picture signaling [16]; 8) intra block motion signaling [17]; 9) reversible variable length coding (RVLC) [15]; 10) resynchronization marker [15]; 11) source-coding-level FEC [18]; and redundant pictures (also known as sync pictures for video redundancy coding) [19].

Seven of the above tools, namely intra MB/picture refresh, slice coding, reference picture identification, RPS, data partitioning, spare picture signaling, and redundant slices/pictures, are also supported by H.264/AVC. In addition to the "old" standard tools, H.264/AVC includes some new standard tools: 1) parameter sets [20]; 2) Flexible MB Order (FMO) [20]; 3) Gradual Decoding Refresh (GDR) [21]; 4) scene information signaling [22]; 5) SP/SI pictures [23]; 6) constrained intra prediction (see below); and 7) reference picture marking repetition (RPMR, see below).

Nonstandard error control tools include error concealment [15], error tracking [24], [25], and multiple description coding (MDC) [26]. Basically, all the nonstandard tools can be used with any video codec, including H.264/AVC and SVC. However, only a subset of MDC methods, e.g., the one reported in [27], generates standard-compatible bit streams.

Among all the above-mentioned standard error resilient coding tools, reference picture identification, spare picture signaling, GDR, scene information signaling, constrained intra prediction, and RPMR have not been covered by the earlier review papers in [2], [15], [20], [23], and are supported by H.264/AVC or SVC. These tools are reviewed in the following section. In addition, intra refresh and redundant slices/pictures are also reviewed, as the former is the basis for the discussion of SVC LA-RDO algorithm in Section IV, and for the latter there have been considerable amount of new developments since the old review in [20]. For nonstandard error control tools, only error concealment is reviewed, to form the basis for the discussions of SVC error concealment methods in Section IV. Readers are referred to the corresponding references listed above for those error resilient tools that are not covered by the following reviews and detailed discussions.

A. Standard Error Resilient Coding Tools in H.264/AVC

In this section, the standard error resilient coding tools in H.264/AVC are summarized.

1) Reference Picture Identification: In H.264/AVC, each reference pictures is with an incremental frame number. This design frame number enables decoders to detect loss of reference pictures and take proper actions when there are losses of reference pictures.

2) Gradual Decoding Refresh (GDR): GDR is enabled by the so-called isolated region technique [21]. An isolated region evolving over time can completely stop error propagation resulting from packet losses occurring before the starting point of the isolated region in a gradual manner, i.e., after the isolated region covers the entire picture area. It can also be used for other purposes such as gradual random access.

3) Redundant Slices/Pictures: Various usages of redundant slices/pictures are proposed in [27]–[29]. Furthermore, H.264/AVC-compatible redundant picture coding in combination with RPS, reference picture list reordering (RPLR), and adaptive redundant picture allocation was reported in [30].

4) Reference Picture Marking Repetition (RPMR): RPMR, using the decoded reference picture marking repetition SEI message, can be used to repeat the decoded reference picture marking syntax structures in the earlier decoded pictures. Consequently, even if earlier reference pictures were lost, the decoder can still maintain correct status of the reference picture buffer and reference picture lists.

5) Spare Picture Signaling: The spare picture SEI message, which signals the similarity between a reference picture and other pictures, tells the decoder which picture can be used as a substituted reference picture or can be used to better conceal the lost reference picture [16].

6) Scene Information Signaling: The scene information SEI message provides a mechanism to select a proper error concealment method for intra pictures, scene-cut pictures, and gradual scene transition pictures at the decoder [22].

7) Constrained Intra Prediction: In the constrained intra prediction mode, samples from inter coded blocks are not used for intra prediction. Consequently, temporal error propagation can be efficiently stopped.

8) Intra MB/Picture Refresh: Intra refresh intentionally inserts intra pictures or intra MBs into the bit stream. It can achieve better RD performance on certain packet loss conditions. Several methods for insertion of intra MBs have been reported, e.g., random intra refresh (RIR) [31], cyclic intra refresh (CIR) [32], recursive optimal per-pixel estimate (ROPE) [33], sub-pixels ROPE [34], LA-RDO algorithm in H.264/AVC [35], and 4×4 block-based error propagation map method [36].

B. Error Concealment for H.264/AVC

Error concealment is a decoder-only technique. Typically, the spatial, temporal, and spectral redundancy can be made use of to mask the effect of channel errors at the decoder.

If the picture is partially corrupted, e.g., the picture is split into multiple slices, spatial error concealment method, e.g., as in [37], can be used. For low bit rate video transmission such as 3G wireless systems, usually one picture is coded into only one packet, and loss of a packet implies that the entire picture must be recovered from the previously decoded pictures. The simplest way to solve this problem is by copying the previously decoded picture to replace the lost one. However, if the sequence is with smooth motion, motion copy [38] can be used to improve the performance.

IV. Error Resilient Coding and Error Concealment Tools for SVC

All the standard error resilient video coding tools supported by H.264/AVC are inherited to SVC. However, data partitioning and SP/SI pictures are not included in the currently specified SVC profiles. All the nonstandard error control tools are supported by SVC, in the same manner as H.264/AVC. Some of these tools that are inherited from H.264/AVC are supported in the SVC reference software, namely the joint scalable video model (JSVM). These tools are briefly summarized in Section IV-A.

Besides the tools inherited from H.264/AVC, SVC includes three new standard error resilient coding tools, namely quality layer integrity check signaling, redundant picture property signaling, and temporal level zero index signaling. These tools are discussed in Section IV-B.

The conventional error resilient coding and error concealment tools for single-layer coding can certainly be applied to the SVC enhancement layers. However, these methods do not utilize the correlations between different layers, which are high in many cases. Improved performance can be expected if inter-layer correlations are utilized. In Sections IV-C and IV-D, we discuss LA-RDO-based intra MB refresh and error concealment algorithms, respectively, that utilize inter-layer correlations in SVC bit streams.

A. Error Control Tools Inherited from H.264/AVC and Supported in the JSVM

The JSVM software include the support of FMO [39], redundant pictures [40], [41], slice coding [42], LA-RDO-based intra MB refresh [43], as well as some error concealment methods [44], [45].

The simplest exact-copy redundant coding for each picture was proposed to the JSVM by [40]. An unequal error protection (UEP) like method, which only codes redundant representations for key pictures of enhancement layers, was proposed in [41]. The LA-RDO-based intra MB refresh algorithm, which was proposed in [43], was extended from the single-layer method reported in [36]. Four error concealment methods were proposed in [44] according to the inter-layer prediction characteristics of SVC. Another improved error concealment method using motion copy for key picture was proposed in [45]. It has also been agreed to include it in the JSVM software, but at the time of writing the feature has not yet been integrated. By applying some of these error concealment methods in a combined manner, significant PSNR gain compared to single layer error concealment algorithms can be observed.

B. New Standard Error Resilient Coding Tools in SVC

1) Quality Layer Integrity Check Signaling: The quality layer integrity check SEI message includes a cyclic redundancy check (CRC) code calculated from all the quality enhancement NAL units (with the syntax element quality_id larger than 0) of a dependency representation (all NAL units in one access unit and with the same value for the syntax element depencency_id). This information can be used to verify whether all quality NAL units of a dependency representation are received by the decoder. If loss is detected, the decoder can inform the loss to the encoder, which in turn decides the use of the error-free base quality layer as reference for encoding subsequent access units. Therefore the drift error by using the erroneous highest quality layer as reference can be avoided. When no loss is detected, the encoder is free to use the highest quality layer as reference for improved coding efficiency. More details can be found in [46].

2) Redundant Picture Property Signaling: The redundant picture property SEI message can be used to indicate the correlations between a redundant layer representation and the corresponding primary layer representation. A layer representation consists of all NAL units in one dependency representation and with the same value for the syntax element quality_id. Indicated information includes, when a primary picture is lost, whether redundant representation can completely replace the primary representation:

1) for inter prediction or inter-layer prediction;
2) for inter-layer mode prediction (part of inter-layer motion prediction);
3) for inter-layer motion prediction;
4) for inter-layer residual prediction;
5) for inter-layer texture prediction.

More details can be found in [41].

3) Temporal Level Zero Index Signaling: The temporal level zero dependency representation index SEI message provides a mechanism to detect whether a dependency representation at the lowest temporal level (i.e., with temporal_id equal to 0) needed for decoding the current access unit is available when NAL unit losses are expected during transport. Decoders can use the SEI message to determine whether to transmit a feedback message or a retransmission request concerning a lost dependency representation at the lowest temporal level. More details can be found in [47]–[49].

C. LA-RDO-Based Intra MB Refresh for SVC

In SVC, when encoding an MB in an enhancement layer picture, the traditional MB coding modes in single-layer coding as well as new inter-layer prediction mode can be used. Similar as in single-layer coding, MB mode selection in SVC also affects the error resilient performance of the encoded bit stream. In the following, a method that is extended from the single-layer method in [36] to multilayer coding is presented. In this method, given the target packet loss rate (PLR), the 4×4 block-based error propagation maps for a picture is calculated, and the map is taken into account to perform mode decision for pictures in the latter.

In order to understand the multilayer method better, we first discuss the generic LA-RDO process and the particular single-layer method in [36].

1) Mode Decision: The MB mode selection is decided according to the following steps.

1) Loop over all the candidate modes, and for each candidate mode, estimate the distortion of the reconstructed MB resulting from both packet losses and source coding, and the coding rate (e.g., the number of bits for representing the MB).
2) Calculate each mode's cost, which is represented by the following equation, and choose the mode that gives the smallest cost

$$C = D + \lambda R. \tag{1}$$

In (1), C denotes the cost, D denotes the estimated distortion, R denotes the estimated coding rate, and λ is the Lagrange multiplier.

2) Single-layer Method: Assume that the PLR is p_l. The overall distortion of the mth MB in the nth picture with the candidate coding option o is represented by

$$\begin{aligned} D(n,m,o) = {} & (1-p_l)(D_s(n,m,o) + D_{ep\_ref}(n,m,o)) \\ & + p_l D_{ec}(n,m) \end{aligned} \tag{2}$$

where $D_s(n,m,o)$ and $D_{ep\_ref}(n,m,o)$ denote the source coding distortion and the error propagation distortion, respectively; and $D_{ec}(n,m)$ denotes the error concealment distortion in case the MB is lost. Obviously, $D_{ec}(n,m)$ is independent of the MBs coding mode. The source coding distortion $D_s(n,m,o)$ is the distortion between the original signal and the error-free reconstructed signal.

Source coding distortion $D_s(n,m,o)$ is the distortion between the original signal and the error-free reconstructed signal. It can be calculated as the mean square error (MSE), sum of absolute difference (SAD), or sum of square error (SSE). The error concealment distortion $D_{ec}(n,m)$ can be calculated as the MSE, SAD, or SSE between the original signal and the error concealed signal. The used norm, i.e., MSE, SAD or SSE, shall be aligned for $D_s(n,m,o)$ and $D_{ec}(n,m)$.

For the calculation of the error propagation distortion $D_{ep\_ref}(n,m,o)$, a distortion map D_{ep} for each picture on a block basis (e.g., 4×4 luminance samples) is defined. Given the distortion map, $D_{ep\_ref}(n,m,o)$ is calculated as

$$\begin{aligned} D_{ep\_ref}(n,m,o) &= \sum_{k=1}^{K} D_{ep\_ref}(n,m,k,o) \\ &= \sum_{k=1}^{K}\sum_{l=1}^{4} w_l D_{ep}(n_l,m_l,k_l,o) \end{aligned} \tag{3}$$

where K is the number of blocks in one MB, and $D_{ep\_ref}(n,m,k,o)$ denotes the error propagation distortion of the kth block in the current MB. $D_{ep\_ref}(n,m,k,o)$ is calculated as the weighted average of the error propagation distortion $\{D_{ep}(n_l,m_l,k_l,o_l)\}$ of the blocks $\{k_l\}$ that are referenced by the current block. The weight w_l of each reference block is proportional to the area that is used for reference.

The distortion map with the optimal coding mode o^* is defined as follows.

For an inter-coded block wherein bi-prediction is not used, i.e., there is only one reference picture used

$$D_{ep}(n,m,k) = (1-p_l)D_{ep\_ref}(n,m,k,o^*) + p_l(D_{ec\_rec}(n,m,k,o^*) + D_{ec\_ep}(n,m,k)) \quad (4)$$

where $D_{ec\_rec}(n,m,k,o^*)$ is the distortion between the error-concealed block and the reconstructed block, and $D_{ec\_ep}(n,m,k)$ is the distortion due to error concealment and the error propagation distortion in the reference picture that is used for error concealment. Equation (3) is used to calculate $D_{ec\_ep}(n,m,k)$ assuming that the error concealment method is known, i.e., $D_{ec\_ep}(n,m,k)$ is calculated as the weighted average of the error propagation distortion of the blocks that are used for concealing the current block, and the weight w_l of each reference block is proportional to the area that is used for error concealment.

For an inter-coded block wherein bi-prediction is used, i.e., there are two reference pictures used

$$\begin{aligned} D_{ep}(n,m,k) = {} & w_{r0} \times ((1-p_l)D_{ep\_ref\_r0}(n,m,k,o^*) \\ & + p_l(D_{ec\_rec}(n,m,k,o^*) + D_{ec\_ep}(n,m,k))) \\ & + w_{r1} \times ((1-p_l)D_{ep\_ref\_r1}(n,m,k,o^*) \\ & + p_l(D_{ec\_rec}(n,m,k,o^*) + D_{ec\_ep}(n,m,k))) \end{aligned} \quad (5)$$

where w_{r0} and w_{r1} are, respectively, the weights of the two reference pictures used for bi-prediction.

For an intra-coded block, no error propagation distortion is transmitted, and only error concealment distortion is considered

$$D_{ep}(n,m,k) = p_l(D_{ec\_rec}(n,m,k,o^*)+D_{ec\_ep}(n,m,k)) \quad (6)$$

According to [50] the error-free Lagrange multiplier is represented by

$$\lambda_{ef} = -\frac{dD_s}{dR}. \quad (7)$$

However, when transmission error exists, a different Lagrange multiplier may be needed.

Combining (1) and (2), we get

$$C = (1-p_l)(D_s(n,m,o) + D_{ep\_ref}(n,m,o)) + p_l D_{ec}(n,m) + \lambda R. \quad (8)$$

Let the derivative of C to R be zero, and we get

$$\lambda = -(1-p_l)\frac{dD_s(n,m,o)}{dR} = (1-p_l)\lambda_{ef}. \quad (9)$$

Consequently, (1) becomes

$$C = (1-p_l)(D_s(n,m,o) + D_{ef\_ref}(n,m,o)) + p_l D_{ec}(n,m) + (1-p_l)\lambda_{ef} R. \quad (10)$$

Since $D_{ec}(n,m)$ is independent of the coding mode, it can be removed. After $D_{ec}(n,m)$ is removed, the common coefficient $(1-p_l)$ can also be removed, which finally results in

$$C = D_s(n,m,o) + D_{ep\_ref}(n,m,o) + \lambda_{ef} R. \quad (11)$$

3) Multilayer Method: In scalable coding with multiple layers, the MB mode decision for the base layer pictures is exactly the same as in the single-layer method. For a slice in an enhancement layer picture, if no inter-layer prediction is used, the single-layer method is used, with the used PLR being the PLR of the current layer. Otherwise (if inter-layer prediction is used), the distortion estimation and the Lagrange multiplier selection processes are presented below.

Let the current layer contain the current MB be l_c, the lower layer contain the co-located MB used for inter-layer prediction by the current MB be $l_{c\text{-}1}$, the further lower layer containing the MB used for inter-layer prediction of the co-located MB in $l_{c\text{-}1}$ be $l_{c\text{-}2}, \ldots$, and the lowest layer containing an inter-layer-dependent block for the current MB be l_0, and let the PLRs be $p_{l,c}, p_{l,c\text{-}1}, \ldots, p_{l,0}$, respectively. For a current slice that may use inter-layer prediction, it is assumed that a contained MB would be decoded only if the MB and all the dependent lower-layer blocks are received; otherwise the slice is concealed. For a slice that does not use inter-layer prediction, a contained MB would be decoded as long as it is received.

The overall distortion of the mth MB in the nth picture in layer l_c with the candidate coding option o is represented by

$$D(n,m,o) = \left(\prod_{i=0}^{c}(1-p_{l,i})\right)(D_s(n,m,o) + D_{ep\_ref}(n,m,o)) + \left(1 - \prod_{i=0}^{c}(1-p_{l,i})\right) D_{ec}(n,m) \quad (12)$$

where $D_s(n,m,o)$ is calculated the same way as in the single-layer method. $D_{ec}(n,m)$ is determined by the chosen error concealment method. Given the distortion map of the reference picture in the same layer or in the lower layer (for inter-layer texture prediction), $D_{ep\_ref}(n,m,o)$ is calculated using (3).

The distortion map is derived as presented in below. When the current layer is of a higher spatial resolution, the distortion map of the lower layer $l_{c\text{-}1}$ is first upsampled. For example, if the resolution is changed by a factor of two for both the width and the height, then each value in the distortion map is simply upsampled to be a 2×2 block of identical values.

1) Texture prediction: In this mode, distortion can be propagated from the lower layer. Then the distortion map of the kth block in the current MB is as in (13). Note that here $D_{ep\_ref}(n,m,k,o^*)$ is the distortion map of the kth block in the co-located MB in the lower layer l_{n-1}. $D_{ec\_rec}(n,m,k,o^*)$ and $D_{ec\_ep}(n,m,k)$ are calculated the same as in the single-layer method

$$\begin{aligned} D_{ep}(n,m,k) = {} & \left(\prod_{i=0}^{c}(1-p_{l,i})\right) D_{ep\_ref}(n,m,k,o^*) \\ & + \left(1 - \prod_{i=0}^{c}(1-p_{l,i})\right) \\ & \times (D_{ec\_rec}(n,m,k,o^*) + D_{ec\_ep}(n,m,k)). \end{aligned} \quad (13)$$

2) Motion prediction: Since the motion prediction in JSVM use the motion vector field, reference indices and MB partitioning of the lower layer are for the corresponding MB in the current layer. The inter prediction process still uses the reference pictures in the same layer. For a block that uses inter-layer motion prediction and does not use bi-prediction, the distortion map of the kth block is

$$D_{ep}(n,m,k) = \left(\prod_{i=0}^{c}(1-p_{l,i})\right) D_{ep\_ref}(n,m,k,o^*) + \left(1-\prod_{i=0}^{c}(1-p_{l,i})\right)(D_{ec\_rec}(n,m,k,o^*) + D_{ec\_ep}(n,m,k)). \quad (14)$$

For a block that uses inter-layer motion prediction and also uses bi-prediction, the distortion map of the kth block is

$$\begin{aligned} D_{ep}(n,m,k) &= w_{r0} \times \left(\left(\left(\prod_{i=0}^{c}(1-p_{l,i})\right) D_{ep\_ref\_r0}(n,m,k,o^*)\right.\right. \\ &+ \left(1-\prod_{i=0}^{c}(1-p_{l,i})\right)(D_{ec\_rec}(n,m,k,o^*) \\ &+ D_{ec\_ep}(n,m,k))) + w_{r1} \times \left(\left(\prod_{i=0}^{c}(1-p_{l,i})\right)\right. \\ &\times D_{ep\_ref\_r1}(n,m,k,o^*) + \left(1-\prod_{i=0}^{c}(1-p_{l,i})\right) \\ &\left.\left.\times (D_{ec\_rec}(n,m,k,o^*) + D_{ec\_ep}(n,m,k)\right)\right). \end{aligned} \quad (15)$$

Note that here $D_{ep\_ref}(n,m,k,o^*)$ in (14) and $D_{ep\_ref\_r0}(n,m,k,o^*)$ and $D_{ep\_ref\_r1}(n,m,k,o^*)$ in (15) are the distortion map of the kth block calculated from reference pictures in the same layer. $D_{ep\_ec}(n,m,k,o^*)$ and $D_{ec\_ep}(n,m,k,o^*)$ are calculated the same as in the single-layer method.

1) Residual prediction: If the low layer is received, and residue of the low layer can be decoded correctly, then there is no error propagation. Otherwise, the error concealment is performed. Therefore, (14) and (15) can also be used to derive the distortion map for an MB mode using inter-layer residual prediction.
2) No inter-layer prediction: For an inter-coded block, (14) and (15) are used to generate the distortion map, while for an intra-coded block

$$D_{ep}(n,m,k) = \left(1-\prod_{i=0}^{c}(1-p_{l,i})\right) \times (D_{ec\_rec}(n,m,k,o^*) + D_{ec\_ep}(n,m,k)). \quad (16)$$

The calculation process of $D_{ep}(n,m,k)$ can be seen from Fig. 2 clearly.

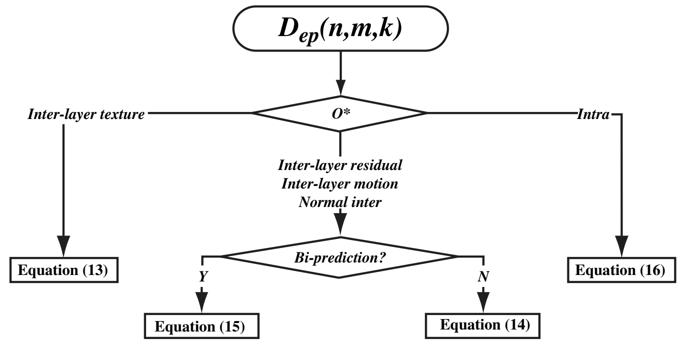


Fig. 2. Calculation of the distortion map $D_{epc}(n,m,k)$.

Combining (1) and (12), we get

$$C = \left(\prod_{i=0}^{c}(1-p_{l,i})\right)(D_s(n,m,o) + D_{ep\_ref}(n,m,o)) \times \left(1-\prod_{i=0}^{c}(1-p_{l,i})\right) D_{ec}(n,m) + \lambda R. \quad (17)$$

Let the derivative of C to R be zero, and then we get

$$\lambda = -\left(\prod_{i=0}^{c}(1-p_{l,i})\right)\left(\frac{dD_s(n,m,o)}{dR}\right) = \left(\prod_{i=0}^{c}(1-p_{l,i})\right)\lambda_{\text{ef}}. \quad (18)$$

Consequently, (1) becomes

$$C = \left(\prod_{i=0}^{c}(1-p_{l,i})\right)(D_s(n,m,o) + D_{ep\_ref}(n,m,o)) \times \left(1-\prod_{i=0}^{c}(1-p_{l,i})\right) D_{ec}(n,m) + \left(\prod_{i=0}^{c}(1-p_{l,i})\right)\lambda_{\text{ef}} R. \quad (19)$$

Here, $D_{ec}(n,m)$ may be dependent on the coding mode, since the MB may be concealed even it is received, while the decoder may utilize the known coding mode to use a better error concealment method. Therefore, the $D_{ec}(n,m)$ term should be retained. Consequently, the coefficient $\prod_{i=0}^{c}(1-p_{l,i})$ that is not common for all the items should also be retained. The final mode decision process becomes

$$C = D_s(n,m,o) + D_{ep\_ref}(n,m,o) + \lambda_{ef} R. \quad (20)$$

Note that the difference between (20) and (11) is that $D_{ep\_ref}(n,m)$ may come from the base layer distortion map if the checked mode o is inter-layer texture prediction and base layer MB is reconstructed. The mode decision process for multilayer is depicted in Fig. 3.

D. Error Concealment Algorithms for SVC

1) Reference Picture Management for Lost Pictures: Upon detection of a lost picture, a key picture is concealed as a lost P picture, and the necessary RPLR commands and

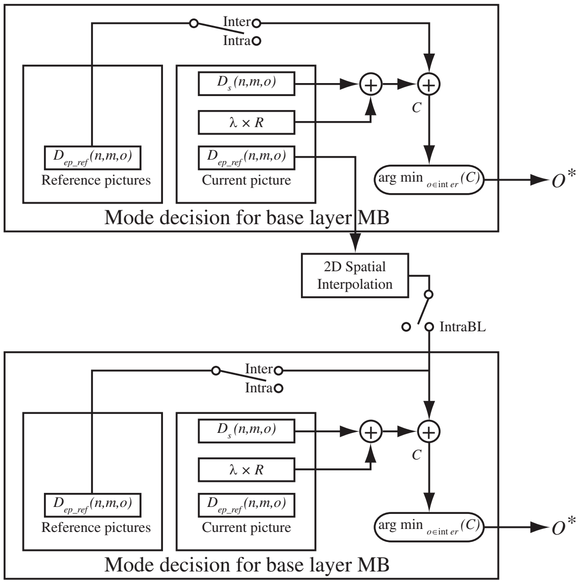


Fig. 3. Mode decision algorithm for the multilayer method.

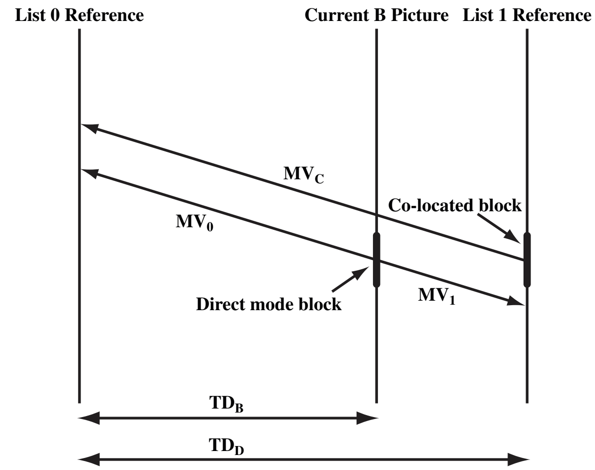


Fig. 4. Example for temporal direct-mode motion vector inference.

memory management control operation (MMCO) commands are set as follows. The RPLR commands are to guarantee the current picture to be predicted from the previous key picture. The MMCO commands are to mark the unnecessary decoded pictures in the previous GOP so as to guarantee the minimum DPB even when packet losses occur. How to conceal a lost key picture is to be discussed in the following sections.

If a lost picture is not a key picture, usually the RPLR commands can be constructed based on those of the pictures in the previous GOPs or on those of the base layer picture if the lost picture is in the enhancement layer.

On the basis of the current design of SVC, the corresponding enhancement layer picture will not be decodable if the base layer picture is lost unless two layers are independently encoded. So base layer picture loss leads to the "loss" of the whole access unit, and one picture of a certain layer leads to the "loss" of the pictures in all the higher layers of the same access unit.

Two types of error concealment algorithms are implemented by us in the current JSVM software. They are summarized as intra-layer error concealment and inter-layer error concealment. One of those methods, if used, is applied to the whole picture, although it is possible that different MBs can selectively use different methods.

2) Intra-layer Error Concealment Algorithms: Intra-layer error concealment is defined as the method that uses the information of the same spatial or quality layer to conceal a lost picture. Three methods are introduced.

1) Picture copy (PC): In this algorithm, each pixel value of the concealed picture is copied from the corresponding pixel of the first picture in the reference picture list 0. If multiple-loop decoding is supported for an error concealment method, this algorithm can be invoked for both the base layer and enhancement layers. Otherwise, only the highest layer in the current access unit can be used for concealment.

2) Temporal direct (TD) for B pictures: The TD mode specified in H.264/AVC is generated as follows. As can be seen in Fig. 4, we assume that an MB or MB partition in the current B picture is coded in temporal direct mode, and then its motion vectors are inferred from its neighboring reference pictures. If the co-located MB or MB partition (belongs to List 1 Reference as shown in Fig. 4) in the reference picture list (namely list for simplicity) one uses a picture (named in Fig. 4 as List 0 Reference) as a reference in list 0 and that picture is also in the list 0 of the current B picture, then the List 0 Reference and List 1 Reference are chosen to bi-predict the being processed MB or MB partition of the current picture. The list 0 and list 1 motion vectors MV_0 and MV_1 are scaled from MV_c using the picture order count (POC, i.e., display order) differences. The detailed deriving process can be seen in [51].

The temporal direct mode specified in H.264/AVC standards cannot be used for any spatial or SNR enhancement layer. However, the concealment of the B picture in SVC can still be applicable for both base layer and enhancement layer. Using the calculated MVs including list 0 and list 1 motion vectors, motion compensation from two specific reference pictures is utilized to predict the MB in the lost picture, assuming zero residue.

In the current SVC design, the necessary motion vectors are stored for each layer. This makes it possible to apply TD at the decoder without introducing extra memory requirement.

1) Motion copy (MC) for key pictures**:** The MC algorithm is applicable for the lost key pictures. Key pictures are concealed as P pictures no mater whether they are originally I or P pictures, since TD is not applicable for this picture and PC may not be efficient because the gap of two key pictures may be large (depending on the GOP size). To get a more accurately concealed picture for the lost key picture, motion vectors are re-generated by copying the motion field of the previous key picture.

3) Inter-layer Error Concealment Algorithms: Two methods are introduced: one works for single-loop decoding; and the other works for multiple-loop decoding.

1) Base layer skip (BLSkip): This method operates as follows. If the base layer is an intra MB, then texture prediction is used. If the base layer is an inter MB, then motion prediction as well as residual prediction are used to generate information for an MB in a lost picture at the enhancement layer. In this case, motion compensation is done at the enhancement layer using the possibly upsampled motion vectors. This algorithm can directly be used for the enhancement layer if there is no picture loss in the base layer. If base layer picture is also lost, the motion vectors for base layer picture are generated using the TD method first. We call this method as BLSkip+TD, but for simplicity we will use BLSkip to represent this method throughout this paper.
2) Reconstruction base layer and possibly upsampling (RU): In the RU algorithm, the base layer picture is reconstructed, and may be upsampled, for the lost picture at the enhancement layer, which is dependent on the spatial ratio between the enhancement layer and the base layer. This requires full decoding of a base layer and thus leads to the requirement of multiple-loop decoding. This method is helpful when there are continuous picture losses only in the enhancement layer and may be competitive for low motion sequences compared with BLSkip.

4) The Improved Error Concealment Algorithm: The improved error concealment method which combines BLSkip with MC is proposed. MC is used to repair the loss of the base layer key picture or those key pictures of the enhancement layer whose base layer pictures are lost. Meanwhile BLSkip is used for the other pictures with losses.

The applicability of these methods is as follows. PC works for all pictures; TD works for all non-key pictures; BLSkip and RU work only for enhancement layer pictures; MC work for key pictures. The RU method can be only used when the decoder adopts multiloop decoding.

V. Simulation

A. Test Conditions

To demonstrate performance of the proposed algorithms, the Bus, Football, Foreman, and News sequences (YUV 4:2:0, 30 frames/s, and progressive) were tested. The tested sequences can be categorized according to their motion characteristics. Bus sequence has high but very regular motion; Foreman sequence has medium but irregular motion; Football sequence has high and irregular motion, while the News sequence has slow motion. The simulation conditions are as follows.

1) JSVM 9.7.
2) Low delay application (IPPP coding structure) and high delay application (hierarchical B picture coding structure with GOP size equal to 16) were tested separately.
3) 4001 pictures were encoded and decoded.
4) Intra picture period: 32 for low delay application and 128 for high delay application.
5) Two layers: base layer was QCIF@30 Hz; enhancement layer was CIF@30 Hz.
6) QP: 28, 32, 36, 40. Base layer and enhancement layer had the same QPs.
7) Multiple slice structure was not used.
8) The error patterns included in [52] are used, and PLRs were as in the following table:

TABLE I
TESTED PLRs

Base layer PLR (%)	0	3	3	5	5	10	10	20
Enhancement layer PLR (%)	3	3	5	5	10	10	20	20

The PLR pair at the encoder for LA-RDO was the same as that of the target PLR pair of the decoder.

The bit stream through packet loss was generated by [53] with two modifications as follows.

1) The base layer was defined as the spatial base layer.
2) Error patterns that determine the packet losses of enhancement layer and base layer packets do not overlap.

The comparisons in the following aspects are considered:

1) with/without LA-RDO;
2) with/without MC.

As it can be concluded from the experimental results of [44] that the BLSkip error concealment method is a good error concealment tool and PC method is preliminary, both of them are considered here as basic algorithms for comparisons. RU requires multiple-loop decoding, thus the results are not reported here but can be found in [44].

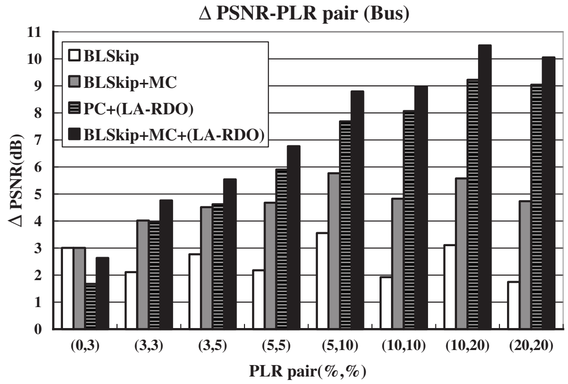
Given different choices, there are various combinations in terms of configurations. However, each of them is compared with the PC case without LA-RDO, which is named as "Anchor" in this section, and the Y-PSNR (luma) differences are calculated by the Bjontegaard measurement [54].

B. Simulation Results for Low-Delay Application

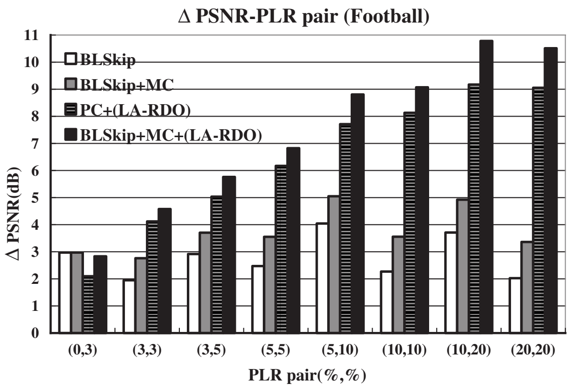
The results are shown in Fig. 5, and we could see that the BLSkip method outperforms the PC method for all tested sequences, with an average PSNR gain of around 2.3 dB over all sequences and all PLR pairs, as summarized in Fig. 6. A further 1 dB gain on average can be achieved by MC, as shown in Fig. 6. LA-RDO provides nearly 5 dB gain on average when PC is utilized. If LA-RDO and BLSkip+MC are combined, an average of more than 5.5 dB gain can be obtained, which outperforms any other methods. However, there may be a few losses in regard to several low PLR pairs compared with "Anchor," which may be caused by some excessive intra MBs introduced by LA-RDO algorithm.

It is also clear that, for low motion sequences (e.g., News), the gains between PC and other advanced error concealment methods are relatively small no matter whether LA-RDO is on or off. Furthermore, the benefits when LA-RDO is off are far from those when LA-RDO is on for the low motion sequence, as shown in Fig. 6.

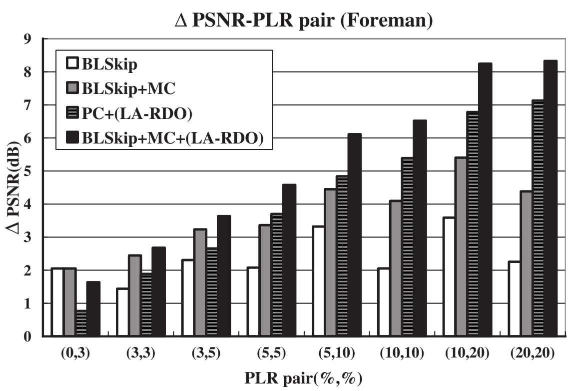
The gains of the above methods, especially when the best configuration is adopted, increase when the PLR pair increases. However, when the PLR pair is very high, e.g.,



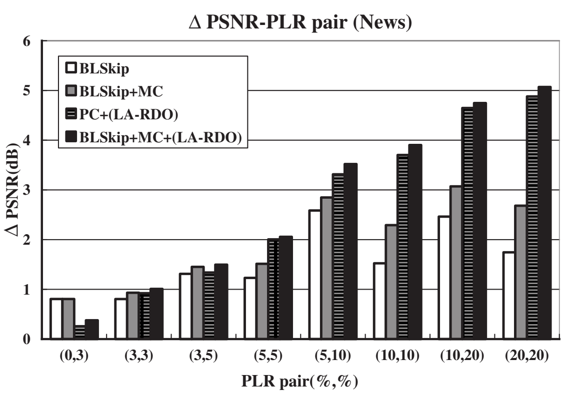
(a)



(b)



(c)



(d)

Fig. 5. △ PSNR (dB) for low delay application.

(20%, 20%), the gains might decrease a little for the high motion sequences (i.e., Bus and Football).

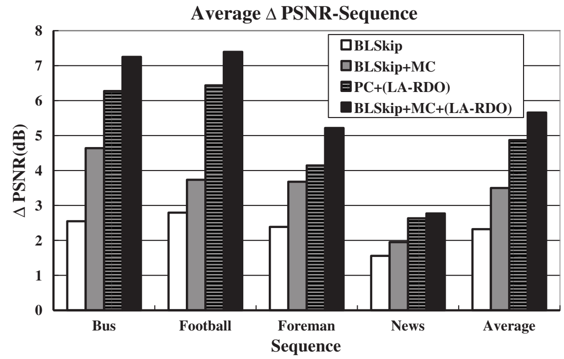


Fig. 6. Average △ PSNR(dB) for low delay application.

Some selective RD curves are plotted in Fig. 7 in order to show the error resilient performance of different methods clearly. It should be noted that, in these figures, if LA-RDO is on, the bit rate will increase much more than in other methods by reason of increasing intra MBs under the same QP setting, so the bit rate ranges for curves with LA-RDO and those for curves without LA-RDO are different. However, there are still some overlapped bit rate ranges, and the trends of these curves are obvious; therefore it is easy to determine which one is the best among different curves.

As can be seen, BLSkip+MC with LA-RDO is the best method among all the methods, while BLSkip+MC is the best when LA-RDO is off.

C. Simulation Results for High-Delay Application

The results are shown in Fig. 8, and the average values are given in Fig. 9. Compared with low-delay application, the results of TD method in high-delay application were provided by extra bars.

From these two figures, we can see that the BLSkip method outperforms the PC method, with an average PSNR gain of around 2.8 dB over all sequences and all PLR pairs. But only a small gain on average can be achieved by MC. LA-RDO provides smaller gains than those of low-delay application, while the average gain is about 2.8 dB when PC is utilized.

Also, the average gain provided by LA-RDO decreases to around 4.0 dB when BLSkip+MC is adopted compared to 5.5 dB in low-delay application. TD method outperforms PC by only about 0.3 dB on average, which is much worse than BLSkip. In conclusion, inter-layer information is of crucial importance for the error concealment algorithm in SVC and is much better than making use of only intralayer information.

Compared to the results shown in Fig. 5 and Fig. 6, most of the observations are still valid, and we skip the detailed analysis of them in this section. However, the differences of the performances in high-delay application are discussed.

The most significant difference is that the average gain for the BLSkip method is higher in high-delay application (about 2.8 dB) than in the low-delay application (about 2.3 dB). The main reason is that in the high-delay application, hierarchical B picture coding structure is used, and therefore the distances from pictures are farther and motion information turns out to be more important. In this case also, PC turns out worse

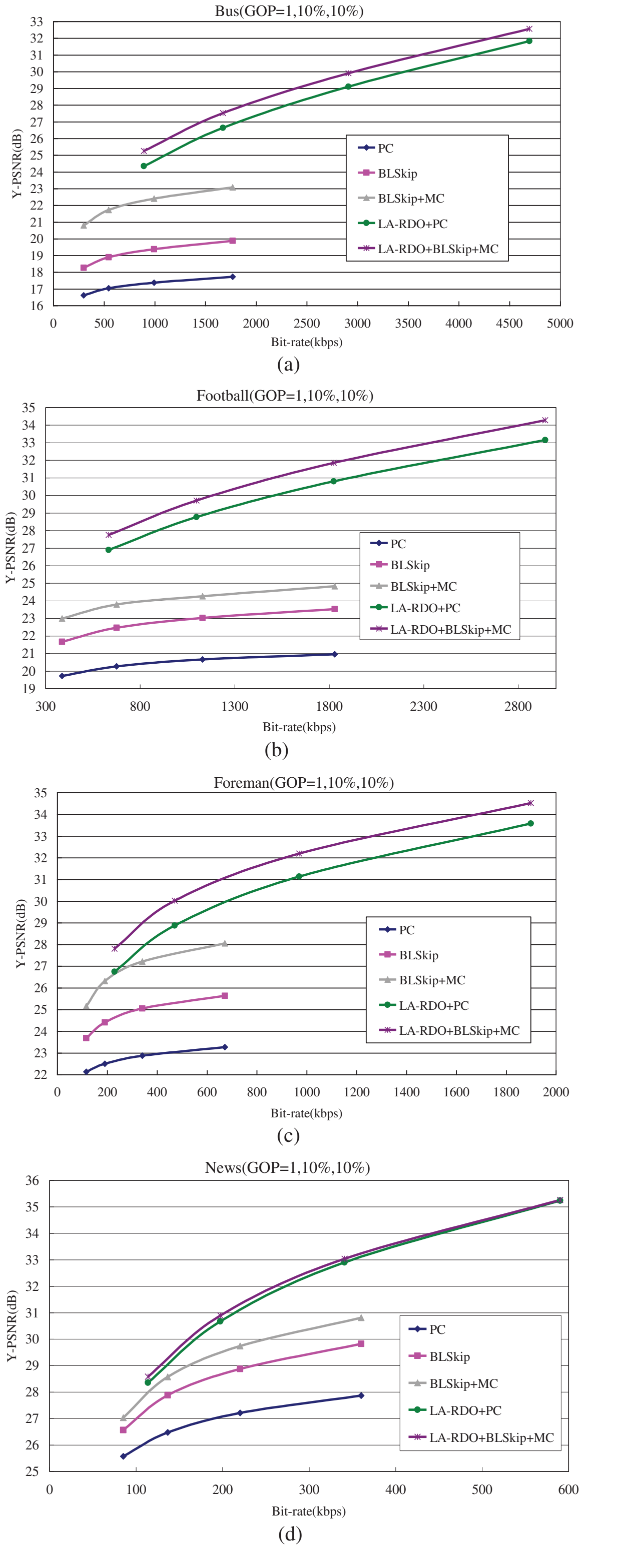


Fig. 7. RD curves of all sequences for the (10%, 10%) PLR pair in low delay application.

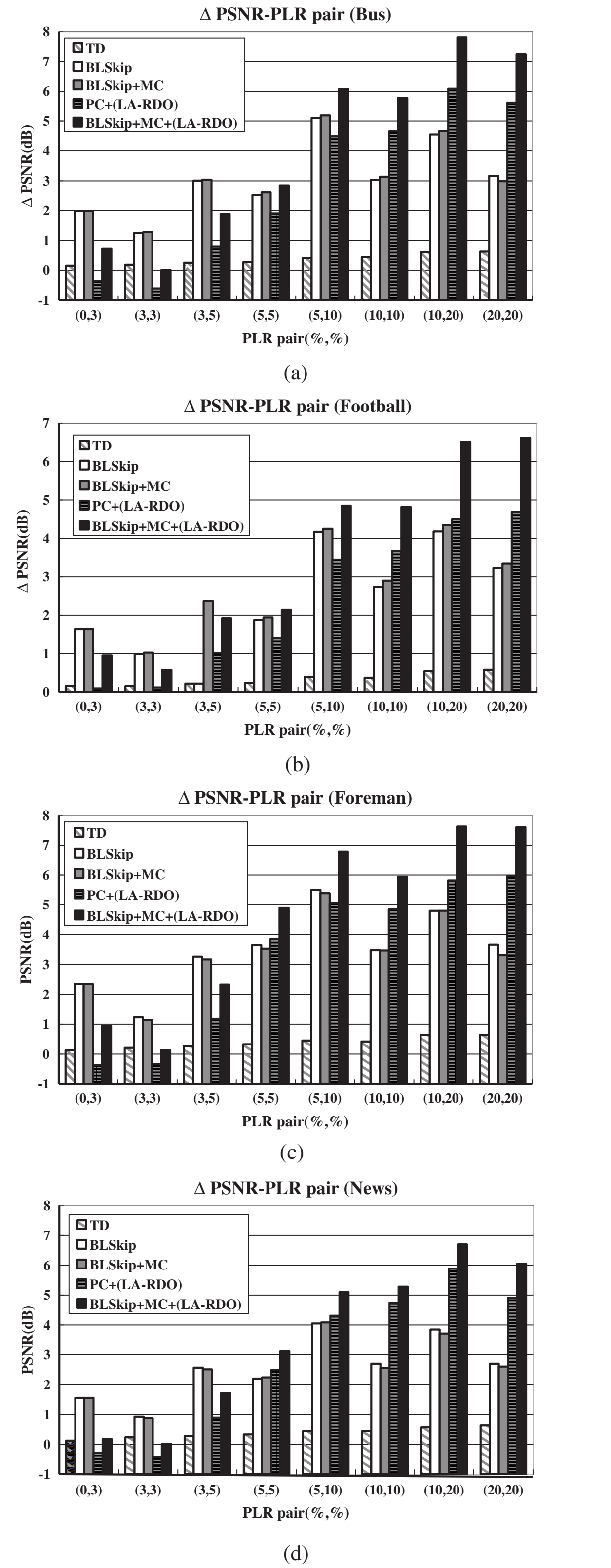


Fig. 8. △ PSNR(dB) for high-delay application.

because those pictures can be used for copying with a larger distance to the lost picture. Temporal motion prediction gets weaker because of the same reason. But this does not affect inter-layer motion prediction used in BLSkip. However, the average gain of Football decreases to about 2.3 dB, which demonstrates that the utilization of inter-layer information will be less effective in the high-delay application for some fast and irregular motion sequences.

The second difference is that the MC method performs much worse, mainly because there is only one key picture in every 16 pictures, and the motion information copied form last

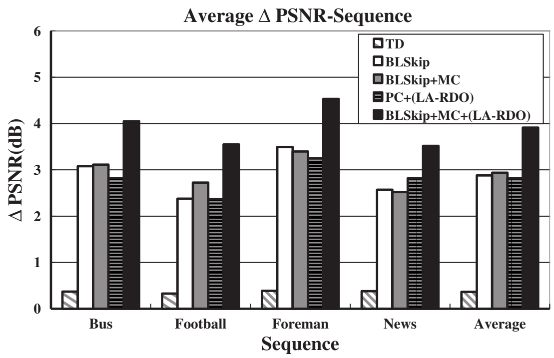


Fig. 9. Average △ PSNR(dB) for high-delay application.

key picture for current key picture will be futile if these two pictures are in different motion speeds and directions. While in low-delay application, every picture is a key picture, and the correlation of motion information between two consecutive pictures is very strong, so considerable gains can be achieved compared with PC method.

The third difference is that the performance of LA-RDO decreases in hierarchical B picture coding structure. In high-delay application, the end-to-end distortion, especially those of B pictures, will be harder to estimate than in low-delay application, and inaccuracy of estimation can sometimes decrease the coding efficiency. In high-delay application, the channel distortion of one B picture may be referred by many other pictures that have higher temporal level, whereas in low-delay application only the latter picture will refer the distortion of the current key picture.

The fourth difference is that, for Bus and Foreman sequences, there are some slight losses which are less than 0.6 dB for several low PLR pairs in high-delay application, at the same time the corresponding gains of these low PLR pairs will be inferior to those gains without LA-RDO. It seems that when LA-RDO is on, there may be some excessive intra MBs in low PLR pairs, which greatly degrade the RD performance. However, for the fast and irregular motion sequence (i.e., Football), the excessive intra MBs can intentionally truncate the channel distortion, so there are no losses in the low PLR pairs.

In summary, BLSkip is much more important in the high-delay application; however, other methods, such as MC and LA-RDO are also helpful, the latter being able to provide about 1 dB extra average gain.

Some selective RD curves are plotted in Figs. 10 and 11 in order to show the error resilient performance of different methods clearly. As can be seen in Fig. 10, BLSkip+MC with LA-RDO is the best method among all the methods, while TD almost gives the same results as those of PC without LA-RDO. In Fig. 11, the RD curves of (3%, 3%) PLR pair is specially given to show that BLSkip or BLSkip+MC without LA-RDO can be suitable for some very low PLR pairs in the high-delay application. The gains are about 1–2 dB compared to other methods.

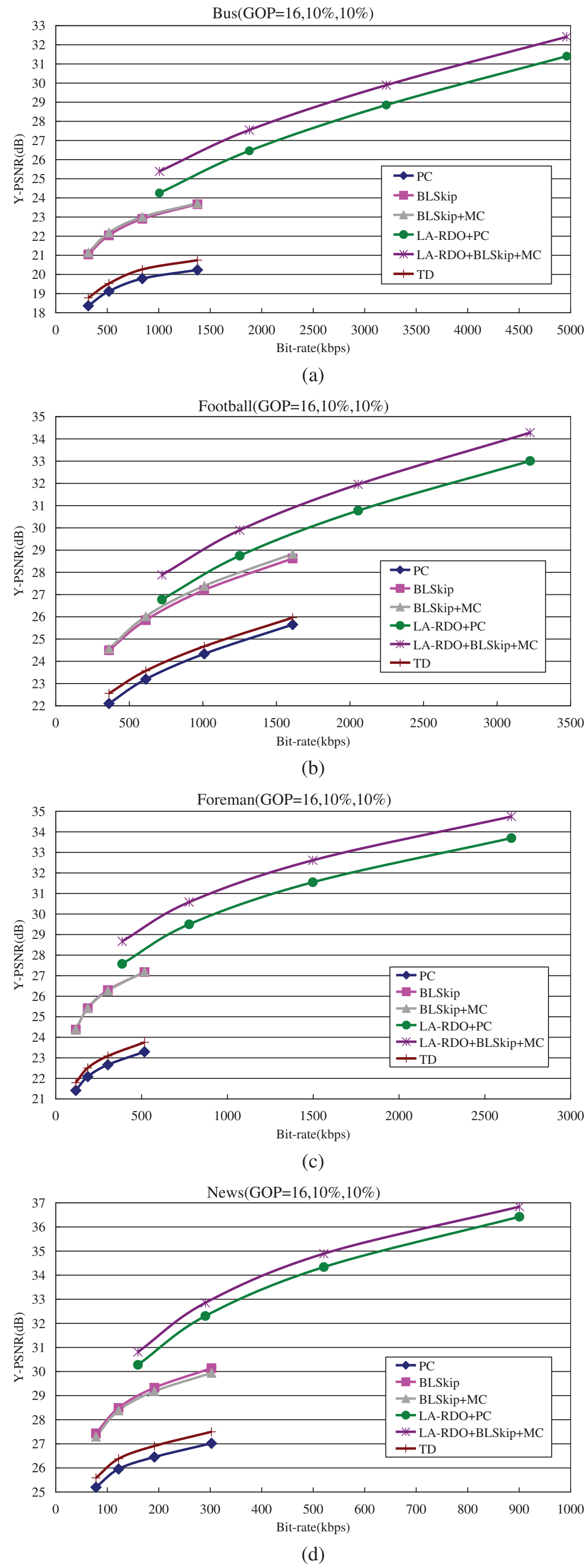


Fig. 10. RD curves of all sequences for (10%, 10%) PLR pair in high-delay application.

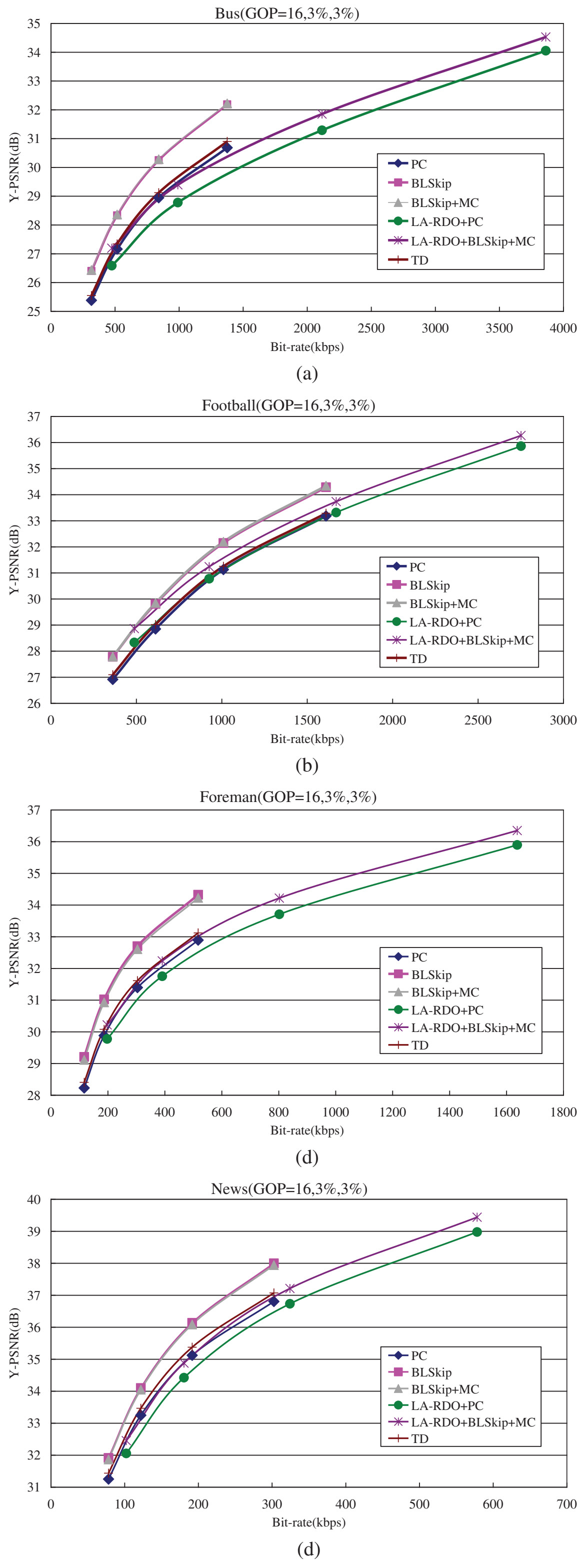


Fig. 11. RD curves of all sequences for (3%, 3%) PLR pair in high-delay application.

VI. Conclusion

SVC has been recently approved as an international standard. Apart from better coding efficiency, it provides improved adaptation capability to heterogeneous network compared to earlier SVC standards. Error resilient coding and error concealment are highly desired for the robustness and flexibility of SVC-based applications. In this paper, we reviewed error resilient coding and error concealment algorithms in H.264/AVC and SVC. LA-RDO algorithm for SVC was presented in detail. Moreover, five error concealment methods for SVC were proposed and analyzed. Simulation results showed that LA-RDO for SVC, the proposed error concealment methods, and their combination improve the average picture quality under erroneous channel conditions when compared to the design applying no error-resilient tools at the encoder and only picture copy error-concealment method at the decoder.

Acknowledgment

The authors thank the experts of ITU-T VCEG, ISO/IEC MPEG, and the Joint Video Team (JVT) for their contributions and Kai Xie, Jill Boyce, Purvin Pandit, and Feng Zhang from Thomson for their contributions to the SVC error concealment methods discussed in this paper.

References

[1] J. Postel, *User Datagram Protocol*, IETF STD 6 (RFC 0768), Aug. 1980.
[2] Y. Wang and Q. Zhu, "Error control and concealment for video communication: A Review," *Proc. IEEE*, vol. 86, no. 5, pp. 974–997, May 1998.
[3] Y. Wang, J. Ostermann, Y.-Q. Zhang, *Video Process. and Commun.* Englewood Cliffs, NJ: Prentice Hall, 2002.
[4] T.Wiegand, G. Sullivan, J. Reichel, H. Schwarz, and M.Wien, *Joint Draft 11 of SVC Amendment*, Joint Video Team, Doc. JVT-X201, Jun.–Jul. 2007.
[5] *Advanced Video Coding Generic Audiovisual Services*, ITU-T Rec.H.264|ISO/IEC IS 14496-10 v3, 2005.
[6] *Generic Coding Moving Pictures and Associated Audio Inform.-Part 2: Video*, ITU-T Rec.H.262|ISO/IEC 13818-2 (MPEG-2 Video), Nov. 1994.
[7] ITU-T Rec. H.263, *Video coding for low bit rate communication*, v3: Nov. 2000.
[8] ISO/IEC 14492-2 (MPEG-4 Visual), *Coding of audio-visual objects-Part 2: Visual*, v3: May 2004.
[9] D. Tian, M. M. Hannuksela, and M. Gabbouj, "Sub-sequence video coding for improved temporal scalability," in *Proc. ISCAS '05*, vol. 6. Kobe, Japan, May 2005, pp. 6074–6077.
[10] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of scalable video coding extension of H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.
[11] H. Schwarz, T. Hinz, D. Marpe, and T. Wiegand, "Constrained inter-layer prediction for single-loop decoding in spatial scalability," in *Proc. ICIP '05*, vol. 2. Genova, Ital, Sep. 2005, pp. II-870–II-3.
[12] P. Amon, T. Rathgen, and D. Singer, "File format for scalable video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1174–1185, Sep. 2007.
[13] S. Wenger, Y.-K. Wang, and T. Schierl, "Transport and signaling of SVC in IP networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1164–1173, Sep. 2007.
[14] Y.-K. Wang, M. M. Hannuksela, S. Pateux, A. Eleftheriadis, and S. Wenger, "System and transport interface of SVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1149–1163, Sep. 2007.
[15] Y. Wang, S. Wenger, J. Wen, and A. K. Katsaggelos, "Error resilient video coding techniques," *IEEE Signal Process. Mag.*, vol. 17, no. 4, pp. 61–82, Jul. 2000.
[16] D. Tian, M. M. Hannuksela, Y.-K. Wang, and M. Gabbouj, "Error resilient video coding techniques using spare pictures," in *Proc. Packet Video Workshop '03*, Nantes, France, Apr. 2003.
[17] S. Cen and P. Cosman, "Comparison of error concealment strategies for MPEG video," in *Proc. IEEE Wireless Commun. Networking Conf. (WCNC)*, vol. 1. New Orleans, LA, Sep. 1999, pp. 329–333.
[18] *Video Coding Low Bit Rate Commun.—Annex H: Forward Error Correction Coded Video Signal*, ITU-T Rec. H.263 Annex H, Feb. 1998.

[19] S. Wenger, "Video redundancy coding in H.263+," in *Proc. Int. Workshop Audio-Visual Services Over Packet Networks*, Sep. 1997.

[20] S. Wenger, "H.264/AVC over IP," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 645–656, Jul. 2003.

[21] M. M. Hannuksela, Y.-K. Wang, and M. Gabbouj, "Isolated regions in video coding," *IEEE Trans. Multimedia*, vol. 6, no. 2, pp. 259–267, Apr. 2004.

[22] Y.-K. Wang, M.M. Hannuksela, K. Caglar, and M. Gabbouj, "Improved error concealment using scene information," in *Proc. 2003 Intern. Workshop Very Low Bitrate Video (VLBV'03),* pp. 283–289, Madrid, Spain, Sep. 2003.

[23] S. Kumar, L. Xu, M. K. Mandal, and S. Panchanathan, "Error resiliency schemes in H.264/AVC standard," *J. Vis. Comm. Image Represent.*, vol. 17, no. 2, pp. 425–450, Apr. 2006.

[24] Y.-K. Wang, C. Zhu, and H. Li, "Error resilient video coding using flexible reference frames," in *Proc. SPIE VCIP '05*, pp. 691–702, Beijing, China, Jul. 2005.

[25] B. Girod and N. Farber, "Feedback-based error control for mobile video transmission," *Proc. IEEE*, vol. 87, no. 10, pp. 1707–1723, Oct. 1999.

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

[27] I. Radulovic, Y.-K. Wang, S. Wenger, A. Hallapuro, M. M. Hannuksela, and P. Frossard, "Multiple description H.264 video coding with redundant pictures," in *Proc. Mobile Video Workshop, ACM Multimedia '07*, pp. 37–42, Augsburg, Germany, Sep. 2007.

[28] Y.-K. Wang, M. M. Hannuksela, and M. Gabbouj, "Error resilient video coding using unequally protected key pictures," in *Proc. 2003 Int. Workshop Very Low Bitrate Video (VLBV '03)*, pp. 290–297, Madrid, Spain, Sep. 2003.

[29] S. Rane, P. Baccichet, and B. Girod, "Modeling and optimization of a systematic lossy error protection system based on H.264/AVC redundant slices," in *Proc. Picture Coding Symp. (PCS '06)*, Beijing, China, Apr. 2006.

[30] C. Zhu, Y.-K. Wang, and H. Li, "Adaptive redundant picture for error resilient video coding," in *Proc. ICIP '07*, vol. 4. San Antonio, TX, Sep. 2007, pp. IV-253–IV-256.

[31] G. Cote and F. Kossentini, "Optimal intra coding of blocks for robust video communication over the Internet," *Signal Process. Image Commun.*, vol. 15, no. 1, pp. 25–34, Sep. 1999.

[32] Q. Zhu and L. Kerofsky, "Joint source coding, transport processing and error concealment for H.323-based packet video," in *Proc. SPIE VCIP'99*, pp. 52–62, San Jose, Jan. 1999.

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

[34] H. Yang and K. Rose, "Recursive end-to-end distortion estimation with model-based cross-correlation approximation," in *Proc. ICIP '03*, vol. 2. Barcelona, Spain, Sep. 2003, pp. III-469–III-72.

[35] T. Stockhammer, D. Kontopodis, and T. Wiegand, "Rate-distortion optimization for JVT/H.26L coding in packet loss environment," in *Proc. Packet Video Workshop '02*, Pittsburgh, PA, Apr. 2002.

[36] Y. Zhang, W. Gao, H. Sun, Q. Huang, and Y. Lu. "Error resilience video coding in H.264 encoder with potential distortion tracking," in *Proc. ICIP '04*, vol. 1. Singapore, Oct. 2004, pp. 163–166.

[37] Y.-K. Wang, M. M. Hannuksela, V. Varsa, A. Hourunranta, and M. Gabbouj, "The error concealment feature in the H.26L test model," in *Proc ICIP '02*, vol. 2. Rochester, NY, Sep. 2002, pp. II-729–II-732.

[38] Z. Wu and J. Boyce, "An error concealment scheme for entire frame losses based on H.264/AVC," in *Proc. ISCAS'06*, pp. 4463–4466, Island of Kos, Greece, May 2006.

[39] T. Bae, T. Thang, D. Kim, Y. Ro, J. Kang, J. Kim, and J. Hong, "FMO implementation in JSVM," Poznan, Porland, Doc. JVT-P043, Jul. 2005.

[40] J. Jia, H. Kim, and H. Choi etc., "Implementation of redundant pictures in JSVM," Sejong Univ. and ETRI, Doc. JVT-Q054, Nice, France, Oct. 2005.

[41] C. He, H. Liu, H. Li, Y.-K. Wang, and M.M. Hannuksela, "Redundant picture for SVC," USTC and Nokia Corporation, Doc. JVT-W049, San Jose, Apr. 2007.

[42] S. Tao, H. Liu, H. Li, and Y.-K. Wang, "SVC slice implementation to JSVM," USTC and Nokia Corporation, Doc. JVT-X046, Geneva, Switzerland, Jun. 2007.

[43] Y. Guo, Y.-K. Wang, and H. Li, "Error resilience mode decision in scalable video coding," in *Proc. ICIP'06*, pp. 2225–2228, Atlanta, Oct. 2006.

[44] Y. Chen, K. Xie, F. Zhang, P. Pandit, and J. Boyce, "Frame loss error concealment for SVC," *Journal of Zhejiang University SCIENCE A, also in Proc. Packet Video Workshop'06*, pp. 677–683, Hangzhou, China, Apr. 2006.

[45] Y. Guo, Y.-K. Wang, and H. Li, "Motion-copy error concealment for key pictures," USTC and Nokia Corporation, Doc. JVT-Y047, Shenzhen, China, Oct. 2007.

[46] Y.-K. Wang and M.M. Hannuksela, "SVC feedback based coding," Nokia Corporation, Doc. JVT-W052, San Jose, Apr. 2007.

[47] A. Eleftheriadis, S. Cipolli, and J. Lennox, "Improved error resilience using frame index in NAL header extension for SVC," Layered Media, Inc., Doc. JVT-V088, Marrakech, Morocco, Jan. 2007.

[48] Y.-K. Wang and M.M. Hannuksela, "On tl0_pic_idx in SVC," Nokia Corporation, Doc. JVT-W050, San Jose, Apr. 2007.

[49] A. Eleftheriadis, S. Cipolli, and J. Lennox, "Improved error resilience using temporal level 0 picture index," Layered Media, Inc., Doc. JVT-W062, San Jose, Apr. 2007.

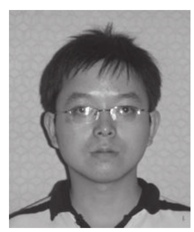
[50] T. Wiegand and B. Girod, "Lagrangian multiplier selection in hybrid video coder control," in *Proc. ICIP '01*, vol. 3. Thessaloniki, Greece, Oct. 2001, pp. 542–545.

[51] M. Flierl and B. Girod, "Generalized B pictures and the draft H.264/AVC video compression standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 587–597, Jul. 2003.

[52] S. Wenger, "Error patterns for Internet experiments," TU Berlin, Doc. VCEG-Q15-I-16r1, New Jersey, Oct. 1999.

[53] Y. Guo, Y.-K. Wang, and H. Li "SVC/AVC loss simulator donation," USTC and Nokia Corporation, Doc. JVT-Q069, Bangkok, Thailand, Jan. 2006.

[54] S. Pateux and J. Jung, "An Excel add-in for computing Bjontegaard metric and additional performance analysis," Orange-France Telecom Research and Development, Doc. VCEG-AE07, Marrakech, Morocco, Jan. 2007.



Yi Guo received the B.S. degree in electronic information engineering from the Department of Electronic Engineering and Information Science at the University of Science and Technology of China, Hefei in 2004. He is currently working toward the Ph.D. degree in signal and information processing at the same university.

During April 2008–June 2008, he was working as an intern at Microsoft Research Asia, Beijing, China. His research interests include image/video processing, image/video coding, error control, and video adaptation.

Ying Chen (M'05) received the B.S. and M.S, degrees in mathematical sciences and electronics engineering and computer science from Peking University, Beijing, China in 2001 and 2004, respectively.

He is currently a Researcher with the Department of Signal Processing at Tampere University of Technology, Tampere, Finland. Before joining Tampere University of Technology, he worked as a Research Engineer at the Thomson Corporate Research, Beijing, China. His research interests include image processing and video coding and transmission. He has been an active contributor to ITU-T JVT and ISO/IEC MPEG, focusing on scalable video coding and multiview video coding standards. He has coauthored over 60 technical standardization reports and over 20 academic papers, and has over 20 issued and pending patents.

Mr. Chen is an external member of Research Staff at the Nokia Research Center, Finland, since September 2006.

Ye-Kui Wang (M'02) received the B.S. degree in industrial automation in 1995 from Beijing Institute of Technology, Beijing, China, and the Ph.D. degree in electrical engineering in 2001 from the Graduate School at Beijing, China, University of Science and Technology of China, Beijing.

From February 2003 to April 2004, he was a Senior Design Engineer at Nokia Mobile Phone. Before joining Nokia, he worked as a Senior Researcher from June 2001 to January 2003 at the Tampere International Center for Signal Processing, Tampere University of Technology, Finland. His research interests include video coding and transport, particularly in an error resilient and scalable manner. He has been an active contributor to different standardization organizations, including ITU-T VCEG, ISO/IEC MPEG, JVT, 3GPP SA4, IETF and AVS. He has been an editor for several (rdraft) standard specifications, including ITU-T Rec. H.271, and the MPEG file format and the IETF RTP payload format for the scalable video coding (SVC) standard. He has also been in the chair of the Special Session of Scalable Video Transport at the 15th International Packet Video Workshop in 2006. He has coauthored over 200 technical standardization contributions and about 40 academic papers. In addition, he has to his credit over 60 issued and pending patents in the fields of multimedia coding, transport, and application systems.

Dr. Wang is currently a Principal Member of the Research Staff with the Department of Signal Processing at Tampere University of Technology, Tampere, Finland.

Miska M. Hannuksela (M'02) received his M.S. and Ph.D. degrees in engineering from Tampere University of Technology, Tampere, Finland, in 1997 and 2009, respectively.

He is currently a Research Leader and the head of the Media Systems and Transport Team in Nokia Research Center, Tampere, Finland. He has more than ten years of experience in video compression and multimedia communication systems. He has been an active delegate in international standardization organizations, such as the Joint Video Team, the Digital Video Broadcasting Project, and the 3rd Generation Partnership Project. His research interests include scalable and error-resilient video coding, real-time multimedia broadcast systems, and human perception of audiovisual quality. He has authored more than 20 international patents and several tens of academic papers.

Houqiang Li received the B.S., M.S. and Ph.D. degrees in 1992, 1997, and 2000, respectively, all in electronic engineering and information science from the University of Science and Technology of China (USTC), Hefei.

From November 2000 to November 2002, he was a Postdoctoral Fellow at the Signal Detection Lab, USTC. Since December 2002, he has been on the faculty and currently he is the Professor with the Department of Electronic Engineering and Information Science at USTC. His current research interests include image and video coding, image processing, and computer vision.

Moncef Gabbouj (M'85–SM'95) received the B.S. degree in electrical engineering in 1985 from Oklahoma State University, Stillwater, and the M.S. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, in 1986 and 1989, respectively.

He is currently a Professor with the Department of Signal Processing at Tampere University of Technology, Tampere, Finland. He was Head of the Department during 2002–2007. His research interests include multimedia content-based analysis, indexing, and retrieval; nonlinear signal and image processing and analysis; and video processing and coding. He is currently on sabbatical leave at the American University of Sharjah, UAE, and Senior Research Fellow of the Academy of Finland.

Dr. Gabbouj has served as Distinguished Lecturer for the IEEE Circuits and Systems Society in 2004–2005. He served as Associate Editor of IEEE TRANSACTIONS ON IMAGE PROCESSING, and was Guest Editor of MULTIMEDIA TOOLS AND APPLICATIONS, the European journal of Applied Signal Processing. He is the past Chairman of the IEEE Finland Section, the IEEE CAS Society, Technical Committee on DSP, and the IEEE SP/CAS Finland Chapter. He was the recipient of the 2005 Nokia Foundation Recognition Award and co-recipient of the Myril B. Reed Best Paper Award from the 32nd Midwest Symposium on Circuits and Systems and co-recipient of the NORSIG 94 Best Paper Award from the 1994 Nordic Signal Processing Symposium.