Error Concealment Techniques for Video Transmission over Error-prone Channels: A Survey

Ziguan CUI\textsuperscript{1,2,*}, Zongliang GAN\textsuperscript{1,2}, Xuefeng ZHAN\textsuperscript{3}, Xiuchang ZHU\textsuperscript{1,2}

\textsuperscript{1}Image Processing and Image Communication Lab, Nanjing University of Posts and Telecommunications, Nanjing 210003, China
\textsuperscript{2}Key Lab of ”Broadband Wireless Communication and Sensor Network Technology”, Nanjing University of Posts and Telecommunications, Ministry of Education
\textsuperscript{3}Institute of Physics and Communication Electronics, Jiangxi Normal University, Nanchang 330027, China

Abstract

Efficient video transmission over unreliable channels may encounter huge challenge due to unavoidable bit error or packets loss. Error concealment (EC) techniques at the decoder side have been developed to recover the damaged regions utilizing spatial or temporal redundant information without changing the encoder structure or adding extra bandwidth. In this work the classic EC techniques and their developments are first reviewed, high-level semantics based EC schemes are also surveyed, and then the emphasis is focused on new EC features introduced by H.264/AVC. Finally, the challenges and future development directions in EC for advanced video coding schemes such as scalable video coding (SVC), multiple description coding (MDC), multi-view video coding (MVC) and stereo video coding are prospected, and future research directions are also indicated according to the current research status and existent problems.

Keywords: Error Concealment; H.264; Video Transmission; Feature-based; Spatio-temporal Adaptive

1 Introduction

Recently, real-time high quality video transmission has gained more and more interests. However, one huge challenge with video transmission over error-prone channels such as wireless or the Internet is that video data packets may be altered or dropped, and some may arrive too late to be useful for decoding and display. Packets loss may bring severe subjective and objective visual quality degradation of the whole group of pictures (GOP) at the decoder, because most video coding standards such as MPEG-x and H.26x are all based on prediction and transform hybrid coding scheme, which makes the encoded bitstream with intensive dependency and very vulnerable to channel error. So the following frames may suffer severe quality degradation due to...
error propagation (EP) from previous damaged ones along the motion compensation path even if they are correctly received, which can be seen clearly in figure 1.

![Error propagation from previous frames without error concealment](image)

Fig. 1: Error propagation from previous frames without error concealment

To combat the annoying effect of packet loss of video transmission over error-prone channels, several error control techniques based on the encoder such as forward error correction (FEC), automatic retransmission request (ARQ) have been developed [1]. Compared with FEC and ARQ, error concealment (EC) techniques acquire extensive research and practical application especially for real-time environment because it does not need extra bandwidth. So this paper focuses on the recent advances in EC technique which is a post-processing scheme at the decoder utilizing the spatial and temporal redundant information in video sequence.

The paper is organized as follows. Section 2 summarizes the classic EC schemes and their developments according to the classification on utilized information, and patch-based and high-level semantics based EC schemes are also surveyed. Section 3 discusses the new EC features introduced by H.264. Section 4 prospects the challenges and future research directions in EC for advanced scalable video coding (SVC), multiple description coding (MDC), multi-view video coding (MVC) and stereo video coding. Section 5 concludes the paper and indicates future research directions.

2 Classic EC Schemes and Their Developments

Based on the information used in EC, this section summarizes the classic and representative EC schemes respectively from frequency, spatial, temporal, spatial-temporal domain combined view. High-level semantics (such as patch and object) based methods are also surveyed. Then the features and applicable environments of different types of EC are compared.

2.1 Frequency domain EC

Frequency domain EC try to recover the DCT coefficients of the lost block by using the coefficients of adjacent blocks, because one common feature within the natural images and video sequences is the smoothness property, which means the DCT coefficients of lost block are likely to be close to the corresponding ones in spatially adjacent blocks.

Frequency Domain EC usually has low complexity and suits for smooth blocks recovery. For example, to obtain smooth image, Alkachouh et al. [2] proposed a simple and efficient method to interpolate each lost pixel only using a selective set of eight border pixels. This method has low computational cost, but the recovery performance is also limited especially for edge and texture blocks because the structure of neighbor blocks is not well exploited and thus blurs the sharp edges within the lost block. Zhai et al. [3] proposed a framework of multiscale DCT pyramid-based Bayesian EC algorithm, which estimated the lost block from DC to AC coefficients gradually.
2.2 Spatial domain EC

To improve the accuracy of pixel recovery, the pixels used for estimation should be spatially close to the lost ones. Spatial EC schemes interpolate the lost pixels directly in pixel domain using the available neighbor blocks. Usually the lost blocks are classified into three types according to the textures inside: flat block, edge block, and texture block, as shown in figure 2.

Various deterministic methods were employed in the previous literature for the interpolation purpose. Wang et al. [4] proposed a simple but effective weighted pixel averaging method which interpolates lost pixels weighted by the inverse distance between the lost pixel and the neighbor border ones. This method can obtain good recovery performance for flat blocks.

As for edge blocks, most methods first estimate the dominant edge directions within the lost block by analyzing the edge information of adjacent blocks and then interpolate the lost pixels along the direction. Sun et al. [5] proposed a spatial EC method based on projection onto convex sets (POCS). The constraint was imposed on the spatial domain to keep the surrounding pixels unchanged, and the most likely edge direction in the neighbor pixels was determined by an edge classifier. This method can obtain good results when the lost block has less dominant edges. However, the POCS method needs an iterative process to get the dominant edge direction, it is very computationally expensive. These methods work well for smooth and edge blocks. But their performance often degrades for the texture complicated blocks. Due to human visual system (HVS) is very sensitive to the reconstructed edges, the challenging problem is that the EC schemes should neither lose true edges nor generate false edges. A considerable research effort has been devoted to the development of edge-preserving interpolation scheme.

To make full use of the local geometric structure information estimated from the surrounding blocks, Zeng et al. [6] proposed a spatial directional interpolation scheme using the statistics of local geometric structure. The method recovers high frequency details well and is computationally efficient for both isolated and contiguous lost blocks. To obtain high detailed content in the lost block reliably, Kim et al. [7], Qaratlu et al. [8] and Ma et al. [9] recover the lost block through fine directional interpolation (FDI). To alleviate the problem of blurring of sharp edges when employing the first-order derivative as the smoothness measure, Zhu et al. [10] proposed to use second-order derivatives as the smoothness measure and preserves the sharp edges effectively. To utilize both neighboring and remote regions information, Wang et al. [11] proposed a best neighborhood matching (BNM) method which uses the block-wise similarities within a frame.

To use recovered pixels in the recovery process afterwards, Li et al. [12] introduced a new sequential recovery framework and adopted an orientation adaptive interpolation derived from the pixelwise statistical model to preserve the edges and high details. The sequential recovery of lost pixels \(X_k (1 \leq k \leq K)\) from available pixels \(Y_l (1 \leq l \leq L)\) can be formulated as follows:

\[
\max p(X_k|X_1, \ldots, X_{k-1}, Y_1, \ldots, Y_L), \text{ for } k = 1, \ldots, K
\]
Zhang et al. [13] proposed a new image restoration method which is virtually different from traditional ways. The major difference is the EC scheme not only uses the information in local areas but also that in remote regions considering there is abundant long-range correlation within natural images and HVS.

2.3 Temporal domain EC

The loss of blocks in inter frames will lose both the motion vector (MV) and residual data. Many temporal EC schemes have been proposed to address the problem by exploiting the temporal correlation between current frame and previously decoded ones. Most algorithms basically involve two steps, first estimate the MV of the lost block, and then recover it using the block pointed by the estimated MV. Usually, temporal EC can obtain better performance than spatial domain methods at regular motion areas.

The simplest method is to set the lost MV as zero and replace the lost macroblock (MB) with co-located MB in the reference frame. It can provide reasonably good estimation for low motion areas such as stationary background, but fail for fast motion and edge areas. Wang et al. [4] proposed the well known boundary matching algorithm (BMA) to select the most suitable MV which shows maximum smoothness between internal and external boundary of the lost MB. To deal with the poor recovery result for the whole block due to incorrect block-based MV estimation, Mualla et al. [14] estimates the MV of each pixel of the lost block using motion field interpolation from the MVs of spatially adjacent blocks and conceals each pixel individually, so incorrect MV estimation will only affect the corresponding pixel.

With increased MV correlation, Zheng et al. [15] use the Lagrange interpolation formula to constitute a polynomial which describes the motion change tendency of neighboring MVs within a small local area, and use this polynomial to recover the lost MV. The format of Lagrange interpolation formula is as follows.

\[ I(x) = y_0L_0^{(n)}(x) + y_1L_1^{(n)}(x) + \cdots + y_nL_n^{(n)}(x) \]  

To improve pixel interpolation performance, Xiang et al. [16] proposed a high efficient temporal EC scheme based on auto-regressive (AR) model which includes a forward AR model for P slice and a bi-direction AR model for B slice. To reduce complexity, Xu et al. [17] proposed a mode switching mechanism to flexibly choose appropriate temporal EC mode for H.264. This method adaptively determines the search range for candidate MVs, and uses a weighted outer boundary distortion function to select the optimal MV for the lost MB. To recover connected corrupted regions, Qian et al. [18] proposed a temporal adaptive EC order determination (AECOD) scheme, which adaptively determines the processing order of an MB by analyzing external boundary patterns of neighboring MBs and deals with the influence of the recovery dependency problem successfully. Seth et al. [19, 20] proposed B-spline approximation based MV recovery method for H.264 to handle non-linear and fast motion. Wang et al. [21] proposed an integrated temporal EC scheme for H.264 with boundary distortion adaptive weight-based switching algorithm. Unlike most of temporal EC, Nguyen et al. [22] did not try to recover the lost MV, but rather based on the sparse representation of image patches on local dictionaries. Gao et al. [23] tried to improve conventional temporal EC from two aspects, one is based on a size-adaptive region basis, and the matching criteria is based on structural similarity (SSIM) instead of MSE or MAD. The method improved the edges recovery effect and more pleasant to HVS due to considering subjective friendly SSIM characteristics.
2.4 Spatio-temporal adaptive EC

Recently, spatio-temporal combined EC schemes [24] have been developed to recover the lost MVs and MBs adaptively utilizing spatial and/or temporal correlation, and thus acquire better recovery results than spatial or temporal EC scheme alone.

To exploit both spatial and temporal correlation, Chen et al. [25] proposed a priority-ranked region matching algorithm, which defines a priority term to determine the restoration order and recovers the lost area region-by-region. Zhang et al. [26] proposed a novel algorithm based on a stochastic modeling. This scheme models the spatial and temporal contextual features in video signals separately using the multiscale Markov Random Field (MMRF), the lost pixels are then estimated using maximum a posteriori (MAP) probabilistic approach. And a new adaptive potential function is introduced to preserve the high frequency information especially the edges.

To reduce the annoying artifacts caused by fast movements, rotations or deformation, Atzori et al. [27] first replaces the lost block with best matching pattern in a previously decoded frame using border information, and then uses an affine transform to deal with the deformable mesh structure. To preserve the structure of lost MB well, Chen et al. [28] proposed a novel two-stage EC scheme. The scheme first reconstructs the lost MV by a novel spatio-temporal boundary matching algorithm (STBMA), and uses a partial differential equation (PDE) based algorithm to refine the pixel recovery. Persson et al. [29] applied Gaussian mixture model (GMM) to EC for block-based packet video. A GMM for video is obtained offline and utilized online to restore lost blocks from spatio-temporal surrounding information. Then extended the above GMM-based method, and proposed a mixture-based estimator and a least squares approach to solve the spatio-temporal combined EC problem [30]. Lately, Zhang et al. [31] obtained two block-dependent AR model coefficients under spatio-temporal continuity constraints, then combined the AR regression results to form the final recovery results. Wu et al. [32] proposed spatio-temporal adaptive EC for H.264 and paid more attention to hardware architecture implement. Cui et al. [33] proposed adaptive EC scheme for H.264 and improved subjective quality especially at scene change or high motions. Zhu et al. [34-37] explored a series of EC methods based on fuzzy reason and proved the effectiveness by extensive experiments.

2.5 Patch-based EC

Recently, patch-based video model has been proposed to implicitly represent temporal motion instead of explicitly block-based motion estimation (ME), which means the motion-related information of pixels has been embedded into the relationship among video patches. A video patch is a generalized image block that can be either 2D or 3D and can have arbitrary shapes. Patch-based method has demonstrated its advantage in image denoising, texture synthesis, regularization, inpainting and EC, and has better restoration performance than traditional methods especially when the damaged area has complicated textures.

Boulanger et al. [38] proposed a novel space-time patch-based method for image sequence restoration, which can adaptively select space-time adjacent patches for each pixel to improve estimation performance without ME. The method can also be combined with ME to deal with large displacements if required, and experiments show that it can improve the quality of highly damaged image sequence greatly. Li et al. [39] proposed a patch-based variational Bayesian framework for video processing. The method first extends block matching into patch clustering
and then exploits the sparsity constraint by sorting and packing similar patches, finally merges diverse inference results from overlapped patches through weighed averaging strategy. Experimental results show that the method obtains dramatic PSNR gain compared with least square based method due to parallel recovery nature which eliminates error propagation effectively.

2.6 Object-based EC

Recently, high-level semantics based techniques have played more and more important role in object based video coding [40] and error control. High-level semantics obtained by abstracting low-level features such as color, texture, shape and motion information can provide better perception characteristics for HVS. However, due to high computational complexity and related techniques immaturity, its application in EC is only limited to shape/texture information recovery in MPEG-4 [41, 42] and content-based image retrieval [43, 44] so far. Using it for practical EC in real-time video transmission still has a long way to go, and the efforts should be directed to reducing the difference of high-level semantics between the original and recovery images and providing better subjective quality. But only a few works concern about this subject until now and still many open issues should be studied.

Shirani et al. [41] employs MAP estimator with an MRF as the image priori model to recover shape information of MPEG-4 video object, and the method restores 20% more missing shape with a better subjective quality compared to the median filtering method. Chen et al. [42] employs the Bezier curve fitting to estimate the intra-video object plane (I-VOP) in spatial domain and uses the weighted side matching criterion to reconstruct the inter-video object planes (P/B-VOP) in temporal domain, and gets better subjective quality than previous works. Lew and Liu et al. [43, 44] surveys the content-based multimedia information retrieval with high-level semantics.

3 New EC Features of H.264/AVC

The latest video coding standard H.264 adopts many new techniques such as multiple modes intra prediction, variable block size inter prediction, and removes the spatial and temporal redundancy greatly, thus achieves higher compression efficiency. The advantages of H.264 make it very suitable for both wired and wireless video applications. But the H.264 coded bitstream is more vulnerable to transmission errors due to the variable length coding (VLC) and other advanced video coding schemes, which makes the EC for H.264 presenting new features. This section summarizes features-based EC techniques and the new challenges introduced by H.264.

3.1 Feature adaptive EC

Due to different EC schemes may have different recovery performance in terms of the feature of lost MB. Several features adaptive EC schemes have been proposed to utilize the advantages of different methods synthetically. The general framework is shown in figure 3.

Agrafiotis et al. [45] proposed a spatial EC scheme that can adaptively switch between directional interpolation (DI) along detected edges and bilinear interpolation (BI) using nearest border pixel based on the directional entropy of detected edge pixels in adjacent received MBs.
Then proposed a scheme which uses the temporal activity measure and edge strength to switch between DI, BI and temporal EC [46].

Based on the coding modes of adjacent MBs, Xu et al. [47] proposed a spatio-temporal adaptive EC scheme with low complexity. The involved temporal EC method uses weighting boundary match to refine subblock-based motion compensation, and improves the ability to deal with high motion areas. The involved spatial EC scheme employs a refined directional weighted spatial interpolation, and protects edge integrity well. Chen et al. [48] proposed a new dynamic multimode switching (DMS) EC scheme for H.264, which provides several EC modes for I and P frames, and adaptively selects suitable EC modes in different situations. These adaptive EC schemes overcome the drawback of existing methods which apply uniform EC mode to entire video sequence regardless of frame features, and thus achieve very good performance gains both in subjective and objective recovery quality.

3.2 EC for whole frame lost

In H.264 low bit rate applications, it is probable to encode a whole frame into one packet considering transmission efficiency. Once bit error occurs in transmission, the whole frame will be lost. The above mentioned methods are all useless because there is no spatial available information in the current frame. In such a scenario, one possible method is to follow the motion tendency of previously decoded frame to give an approximation of the lost frame.

Wu et al. [49, 50] proposed frame copy and MV copy algorithms to conceal the whole lost frame. Frame copy method directly conceals each pixel of the lost frame from the corresponding pixel of previous frame. MV copy method has two steps, first the MVs and reference indices of the co-located blocks in previous reference frame are copied to the lost frame. Then motion compensation is called to reconstruct the lost frame based on the copied MV. These methods can conceal the lost reference or nonreference frame, and even the IDR frames except the first frame in the bitstream. Frame copy algorithm works well in temporal stationary areas but fails in motion areas, because the successive frames which reference the concealed frame will bring error and the error will propagate to the following frames. While MV copy algorithm can alleviate the effect of error propagation greatly. These schemes have been adopted as a non-normative decoder option to the JM reference software due to low implementation overhead. Chien et al. [51] improved MV copy EC method by recursive MV difference refinement process.

Belfiore et al. [52] proposed a whole frame EC scheme based on multiframe optical flow estimation, which includes the following operations: estimation of optical flow for each pixel, spatial regularization of the forward MV field, projection of the last frame onto the missing one, interpolation of missing pixels, last filtering and downsampling. It has better recovery result, but the high computational complexity makes it unsuitable for real time video applications. Bac-
cichet et al. [53] proposed an improved algorithm working at block level instead of pixel level, which presents a little lower performance than Belfiore’s, but its computational simplicity makes it preferable for real time application. The flowchart is shown in figure 4.

![Flowchart](image)

**Fig. 4:** Block layer optical flow based EC flowchart

Hwang et al. [54] first extrapolates the MV of each 44 block of the lost frame by linear motion trajectory feature in the forward and backward prediction directions, and adaptively estimates inter-modes of all MBs in the lost frame using the extrapolated MVs and coding features of H.264. The method has better recovery result than frame copy method and Baccichet’s block level optical flow estimation method, while the computational complexity is close to [53]. Yan et al. [55] proposed an effective method to discard wrongly extrapolated MV and thus improved the results greatly in large motion areas.

### 4 EC for Advanced Video Coding Schemes

Recently, several network-oriented coding schemes such as SVC, MDC, and advanced MVC and stereo video coding have appeared. These advanced coding schemes attract more and more attention due to their flexibility and probably prevail in the near future. Several works have been done to deal with the EC for these advanced coding schemes.

#### 4.1 EC for SVC

SVC as the scalable extension of H.264 can provide flexible adaptability in terms of spatial, temporal and SNR scalabilities for the generated bitstream, and can provide graceful degradation of video quality when transmitted over time-variant lossy channels.

Chen et al. [56] proposed a frame loss concealment scheme for SVC, which proposed four EC methods including frame copy, temporal direct mode motion generation, BISkip mode and reconstruction base layer upsampling. Ji et al. [57] proposed an EC algorithm to tackle the whole frame loss for SVC when hierarchical B-picture coding is used to provide temporal scalability. The scheme derives the motion information of the lost frame simply and efficiently based on the principle of temporal direct mode by the temporal correlation among the adjacent video pictures. Then the lost frame is concealed by performing motion compensation on the correctly received temporally previous and future frames. Experimentation show better concealment result compared with the EC scheme adopted in SVC reference software.

#### 4.2 EC for MDC

MDC is an effective error resilient video coding technique, which can provide adaptability to bandwidth variations and receiving device features. MDC can generate multiple equally important compressed bitstreams (descriptions) for the same video source, which are sent to the decoder side through physically or logically different channels.
To improve the recovery performance of existing EC methods for MDC, Ma et al. [58] proposed a novel multihypothesis error concealment (MHC) algorithm by improving the error recovery rate. The proposed MHC applies temporal interpolation to several additional frames after the lost frame, while existing EC methods only apply concealment to the lost frame.

4.3 EC for MVC and stereo VC

A multi-view video sequence provides multiple views of the same scene, thus it can offer better interactivity and richer user experience. But multi-view video sequences usually contain higher redundancy among different views and spatio-temporal redundancy within a view. MVC is used to efficiently compress multi-view sequence by exploiting the inter-view and spatio-temporal correlations. Once an error occurs in MVC bitstream over the error-prone channel, it will propagate to adjacent views as well as the subsequent frames and degrade the video quality severely.

To acquire high recovery quality even in severe error conditions, Chung et al. [59, 60] proposed an EC algorithm for MVC using four elementary modes to conceal a lost block by combining the best candidate blocks in the temporally adjacent frames and the inter-view frames at the same time instance. To deal with the challenging problem of adaptively combining inter-view and temporal correlation, Xiang et al. [61] proposed an effective hybrid EC method for two-view based 3D video coding. The method first recovers the lost motion or disparity vectors from the MV of neighboring MBs, and then reconstructs the lost MB based on overlapped block motion and disparity compensation, whose weights are determined by the side match criterion and viewpoints.

5 Conclusion

This work first surveys the importance of EC schemes in video transmission over error-prone channels briefly, and summarizes four categories of state-of-the-art EC schemes which are frequency domain EC, spatial EC, temporal EC, and spatio-temporal combined EC. Advanced patch-based and object-based EC schemes are also discussed. Then the new features in EC schemes introduced by H.264 are specially analyzed. The paper also surveys the challenges and future development directions in EC schemes for advanced video coding techniques such as SVC, MDC, MVC and stereo video coding.

According to the current research status and existing problems of EC schemes, the following research issues should be paid more attention: 1) the novel spatial EC methods for texture complicated MBs in I frames from the view of signal processing, which should preserve the true edges while not generate false edges; 2) the spatio-temporal adaptive EC method for H.264 depending on the features of the lost and surrounding MBs; 3) the EC scheme for continuous blocks loss, which is difficult to conceal because the available information is less; 4) high-level semantics based such as patch-based and object-based EC schemes; 5) the whole frame lost EC scheme for low bit rate video transmission especially in the severe error situations such as continuous frames loss, IDR frame loss, the first P frame loss immediately after I frame; 6) the EC schemes for advanced video coding techniques such as SVC, MDC, MVC and stereo video coding because there are only a few literature concerning about these situations.
Acknowledgement

This work is supported by the National Nature Science Foundation of China (No. 61071091, 61071166, 61101105), the priority academic program development of Jiangsu Higher Education Institutions - Information and Communication Engineering, and specialized research fund for the doctoral program of higher education (20113223120001).

References


