Feedback-Based Error Control for Mobile Video Transmission

BERND GIROD, FELLOW, IEEE, AND NIKO FÄRBER, MEMBER, IEEE

Invited Paper

We review feedback-based low bit-rate video coding techniques for robust transmission in mobile multimedia networks. For error control on the source coding level, each decoder has to make provisions for error detection, resynchronization, and error concealment, and we review techniques suitable for that purpose. Further, techniques are discussed for intelligent processing of acknowledgment information by the coding control to adapt the source coder to the channel. We review and compare error tracking, error confinement, and reference picture selection techniques for channel-adaptive source coding. For comparison of these techniques, a system for transmitting low bit-rate video over a wireless channel is presented and the performance is evaluated for a range of transmission conditions. We also show how feedback-based source coding can be employed in conjunction with precompressed video stored on a media server. The techniques discussed are applicable to a wide variety of interframe video schemes, including various video coding standards. Several of the techniques have been incorporated into the H.263 video compression standard recently, and this standard is used as an example throughout.

Keywords**—***Error correction, feedback, mobile communication, standards, video coding.*

I. INTRODUCTION

Of all modalities desirable for future mobile multimedia systems, motion video is the most demanding in terms of bit rate. Even with state-of-the-art compression, television quality requires a few megabits per second (Mb/s), while for low-resolution, limited-motion video sequences typical for picturephones, a few tens of kilobits per second (kb/s) are required for satisfactory picture quality. Today's "second-generation" cellular telephony networks, such as the global system for mobile communications (GSM), typically provide 10–15 kb/s, suitable for compressed speech, but too little for motion video. Fortunately, the standardization of higher bandwidth networks, such as the universal mobile telecommunications system (UMTS) [1], [2], is well underway, and, together with continued progress in video compression technology, mobile video-communicators with picturephone functionality and Internet videoserver access will be possible.

Beyond the limited available bit rate, wireless video transmission offers a number of interesting technical challenges. A recent review has appeared in [3]. One of the issues is that mobile networks cannot provide guaranteed quality of service because high bit error rates occur during fading periods. Transmission errors in a mobile channel range from single bit errors to burst errors or even intermittent loss of the connection. These widely varying error conditions limit the effectiveness of classic forward error correction (FEC), since a worst-case design would lead to a prohibitive amount of redundancy. Closed-loop error control techniques like automatic repeat request (ARQ) [4] have been shown to be more effective than FEC and successfully applied to wireless video transmission [5], [6]. Retransmission of corrupted data frames, however, introduces additional delay, which might be unacceptable for real-time conversational services. Refinements of ARQ schemes proposed for video include the retransmission of more strongly compressed video [7] or the retransmission of multiple copies [8] in a packet network. Both techniques can be combined. Nevertheless, residual transmission errors cannot be avoided with a mobile radio channel, even when FEC and ARQ are combined.

The compressed video signal is extremely vulnerable against transmission errors, since low bit-rate video coding schemes rely on interframe coding for high coding efficiency. They use the previous encoded and reconstructed video frame to predict the next frame. Therefore, the loss of information in one frame has considerable impact on the quality of the following frames. Since some residual transmission errors inevitably corrupt the video bit stream, this vulnerability precludes the use of low bit-rate video coding schemes designed for error-free channels without special measures. These measures have to be built into the video coding and decoding algorithms themselves and form the "last line of defense" if techniques like FEC and ARQ fail. In this paper, we discuss such last-line-of-defense

Manuscript received January 27, 1998; revised March 11, 1998.

The authors are with the Telecommunications Laboratory, University of Erlangen-Nuremberg, Erlangen D-91058 Germany.

Publisher Item Identifier S 0018-9219(99)07472-1.

techniques that can be used to make low bit-rate video coders error resilient. We concentrate on techniques that use acknowledgment information provided by a feedback channel.

The use of a feedback channel is not the only approach to increase the robustness of low bit-rate video communications. A comprehensive review of the great variety of error control and concealment techniques that have been proposed during the last 10–15 years has been presented in an excellent paper by Wang and Zhu recently [8]. For example, one can partition the bit stream into classes of different error sensitivity (often referred to as data partitioning) to enable the use of unequal error protection [9]–[12]. Data partitioning has been included as an error resilience tool in the MPEG-4 standard [13]. Unequal error protection can significantly increase the robustness of the transmission and provide graceful degradation of the picture quality in case of a deteriorating channel. Since unequal error protection does not incorporate information about the current state of the mobile channel, the design of such a scheme is a compromise that accommodates a range of operating conditions. Feedback-based techniques, on the other hand, can adjust to the varying transmission conditions rapidly and make more effective use of the channel.

In this paper, we first review the principles of low bit-rate video coding briefly (Section II), emphasizing the aspects relevant for error-resilience techniques. In Section III, we discuss problems that arise when feeding an erroneous bit stream to an interframe video decoder. In particular, we analyze interframe error propagation. In Section IV, we then present techniques for the video coder to process acknowledgment information from a feedback channel. In Section V, we present simulation results for the transmission of video over a wireless channel to illustrate the effectiveness of feedback-based error control. Finally, in Section VI, we address feedback-based error control for streaming compressed video off a server.

The techniques discussed in this paper are relevant and applicable to a wide variety of interframe video coding schemes, both standard and nonstandard. We emphasize general principles where appropriate. As an illustrative example, we use the video compression standard H.263 [14] throughout. H.263 is not only a state-of-the-art low bit-rate video coder, but it has also been extended to include a variety of feedback-based error control mechanisms. Therefore, this article simultaneously provides an overview of error control techniques currently available in H.263.

II. MOTION-COMPENSATED HYBRID CODING

A. General Principles

Most state-of-the-art low bit-rate video codecs are motion-compensated hybrid codecs, as illustrated in Fig. 1. Two basic modes of operation can be selected, depending on the position of the switch S. These two modes allow the video signal in the current frame to be encoded either directly (INTRA coding), or with reference to previously encoded and reconstructed frames (INTER coding).

The INTER mode combines differential pulse code modulation (DPCM) along an estimated motion trajectory with intraframe encoding of the residual prediction error. Motion-compensated prediction is carried out by estimating the motion between successive frames, shifting the contents of a previously encoded, reconstructed frame accordingly, and transmitting the motion vector in addition to the prediction error residual as side information. Note that the output from the frame memory in Fig. 1 is identical to the decoded frames at the decoder for error-free transmission. Therefore, the same prediction can be formed at the encoder and decoder.

The residual prediction error is usually small and requires fewer bits than directly encoding the original video signal. For efficient encoding, a wide variety of intraframe coding schemes can be used, e.g., subband coding or blockwise vector quantization. In all current compression standards, the discrete cosine transform (DCT) is employed for this purpose with a blocksize of 8 × 8 pixels. The transform coefficients are quantized (Q) and typically encoded as a series of zero-runs and quantizer levels. Transform coefficients and motion vectors are entropy coded along with other side information resulting in variable-length code words, which are multiplexed to the video bit stream.

For most of a video signal, the INTER mode is the preferred mode because of its superior coding efficiency. However, some changes in successive frames, for example, due to uncovered background, new objects appearing in the scene, or after a scene cut, cannot be predicted well, and subtracting the prediction might lead to a prediction error that requires more bits than the original video signal. Therefore, the second basic encoding mode besides INTER coding is the INTRA mode, in which no reference to previous frames is made and the picture is directly intraframe coded. Again, a variety of schemes can be used, but typically a blockwise 8 × 8 DCT coder is employed.

The video codec shown in Fig. 1 is a forward-adaptive system. Decisions about INTER or INTRA mode, motion vector, or quantizer step size are made by the “intelligent” encoder and are transmitted as side information to the “dumb” decoder. This architecture not only reduces the complexity of the decoder but also leads to increased robustness in case of transmission errors, compared to backward-adaptive systems that avoid sending side information by deriving it from the decoded past available both at the encoder and the decoder. Moreover, the forward-adaptive system architecture provides great freedom for the optimization of the coding control since, unlike a backward-adaptive system, the same decoder can be used for a variety of control strategies. In particular, the switching between INTRA and INTER mode is not subject to a prior agreement between encoder and decoder. Feedback-based error control, discussed in Section IV, can exploit this feature. For more information on the general principles of digital video processing, the reader is referred to [15].

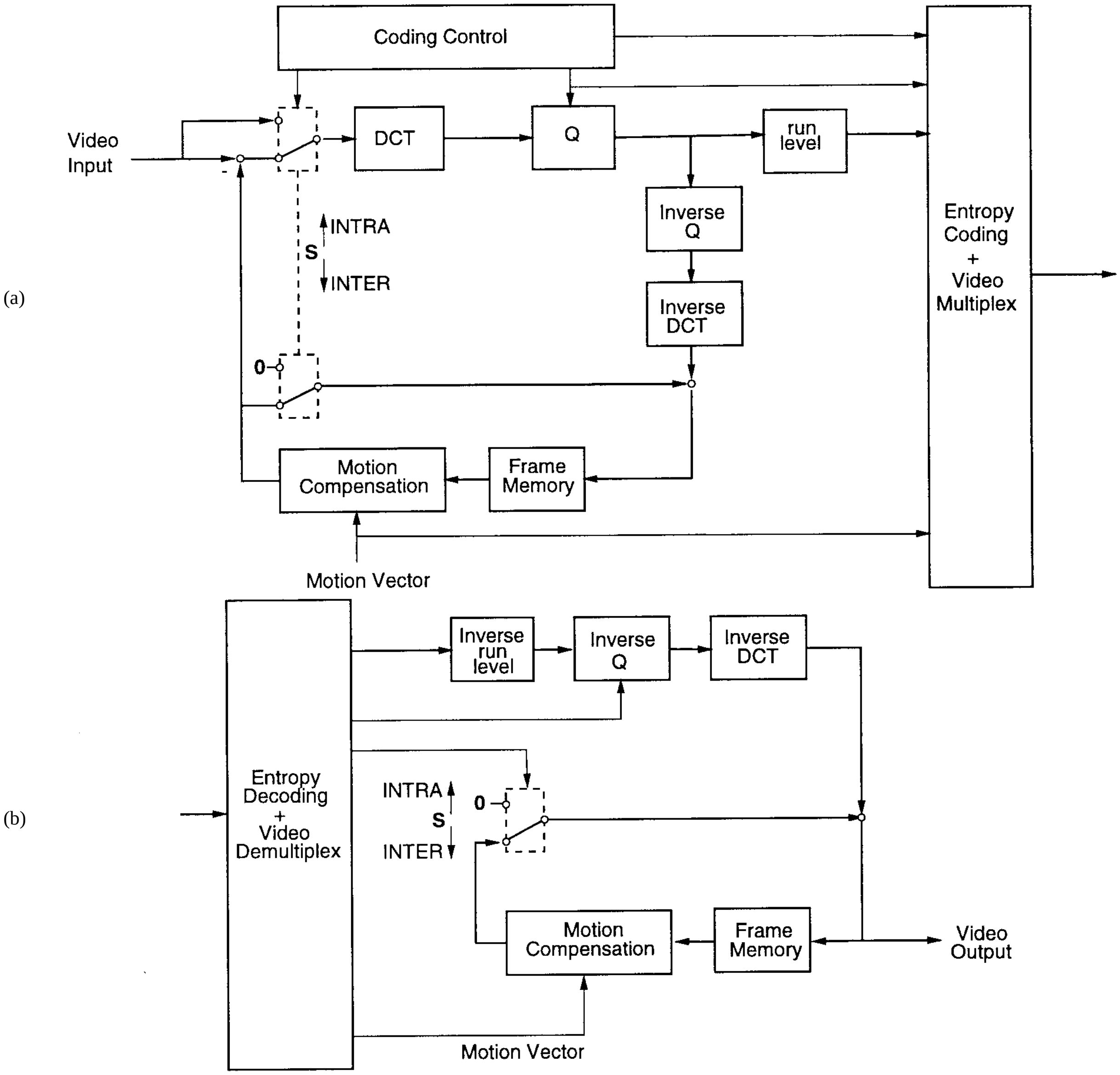


Fig. 1. (a) Motion-compensated hybrid encoder and (b) decoder.

B. The H.263 Coding Standard

The general principles discussed above are the basis for all video compression standards in use today, in particular the ISO standards MPEG-1 [16], MPEG-2 [17], and MPEG-4 [18], and the ITU-T Recommendations H.261 [19], H.262 (identical with MPEG-2), and H.263 [14]. We will use H.263 as an example throughout this paper. For video sequences with moderate motion, the H.263 algorithm can provide compression of at most 100 : 1 to 200 : 1 with acceptable picture quality [20]. To achieve even higher compression, as required for the transmission over mobile networks at very low bit rates, both the spatial resolution and the frame rate have to be reduced compared to standard television pictures. At bit rates below 64 kb/s, frame rates anywhere from 15 f/s (frames per second) down to 5 f/s are common. The picture resolution is often quarter common intermediate format (QCIF, 176 × 144 pixels), which is the most common input format at such low bit rates.

At QCIF resolution, each picture is divided into 11 × 9 macroblocks (MB's), which comprise 16 × 16 luminance samples, and two corresponding 8 × 8 blocks of chrominance samples. The luminance component of each MB is further subdivided into four 8 × 8 blocks, such that 8 × 8 DCT's can be applied to each block. Each MB is encoded either in INTRA or INTER mode. Motion compensation is carried out with half-pixel accuracy with one motion vector sent for each MB. Optionally, individual motion vectors can be provided for each 8 × 8 block, and blocks can be overlapped for motion compensation. A fixed number of successive MB's is usually grouped into a group of blocks (GOB) and side information that is appropriate for a larger number of MB's, but not for an entire frame, can be communicated efficiently on that level. In a similar but more general way, a variable number of MB's can also be grouped into slices.

ITU-T Recommendation H.263 is the video portion of a series of recommendations under the umbrella of ITU-T Recommendation H.324 [21], [22]. H.324 describes terminals for low bit-rate multimedia communication, which may support real-time voice, data, and video, or any combination. Because the transmission is based on V.34 modems operating over the widely available public switched telephone network (PSTN), H.324 terminals are likely to play a major role in future multimedia applications. An increasing number of H.324 terminals is commercially available today. One important reason for this success is the state-of-the-art performance of the H.263 video coding standard [20] that achieves acceptable image quality at modem bit-rates. Other recommendations in the H.324 series include the H.223 multiplex [23], [24], H.245 control [25], and G.723 speech codec [26].

Standardization efforts for mobile H.324 terminals have already started [24]. Like cellular voice telephony, a major requirement is the ability to interwork with terminals connected to the PSTN at a reasonable complexity and with low delay. This precludes the transcoding architecture used today for cellular voice telephony, where a special speech codec is employed only for the mobile part of the network. For low bit-rate video, the delay is typically a few hundred milliseconds due to buffering for constant bit-rate transmission and processing delay. Transcoders would increase this delay unacceptably, and hence end-to-end error control, compatible with H.263, offers significant advantages at lower complexity. Some techniques recently incorporated in H.263 are feedback-based and will be discussed in Section IV.

III. DECODING THE ERRONEOUS VIDEO BIT STREAM

In general, an erroneous bit stream cannot be gracefully decoded by an "off-the-shelf" video decoder build for error-free transmission. Special provisions for error detection, resynchronization, and concealment are required. We discuss these measures and the interframe propagation of the remaining picture impairment in this section.

A. Error Detection and Resynchronization

Because the multiplexed video bit stream consists of variable length code (VLC) words, a single bit error may cause a loss of synchronization and a series of erroneous code words at the decoder. Residual redundancy in non-compact VLC's can be used to design self-synchronizing codes, such that valid symbols may be obtained again after some slippage [27]. However, even if resynchronization is regained quickly, the appropriate location of the decoded information within the frame is no longer known, since the number of missing symbols cannot be determined. Moreover, the subsequent code words are useless if the information is encoded differentially, as it is often the case, e.g., for motion vectors. The common solution to this problem is to insert unique synchronization code words into the bit stream in regular intervals, usually followed by a block of "header" bits. Since any conditional encoding across the resynchronization point must be avoided, the header provides anchor values, e.g., for absolute location in the image or current quantizer step size. Although the length of the synchronization code word can be minimized [28], relatively long synchronization code words are used in practice to reduce the probability of accidental emulation of synchronization words.

As an example, we again consider H.263, which supports optional GOB-headers as resynchronization points. In QCIF format, a GOB consists of 11 MB's that are arranged in one row. Because all information within a correctly decoded GOB can be used independently from previous information in the same frame, the GOB is often used as the basic unit for decoding. Typically, if a transmission error is detected, the GOB is discarded entirely.

Transmission errors can be detected in a variety of ways. If FEC is used, errors can often be detected with high reliability, even if the correction capability of the code is exceeded. For example, in H.261 and H.263 an optional FEC framing can be used to detect errors within the 493 information bits of each block. Reliability information can even be obtained for each received bit when the receiver provides channel state information or a soft output Viterbi algorithm (SOVA) is used for decoding of convolutional codes [29]. This information is then passed on to the video decoder. In addition, the video decoder itself can detect transmission errors. The video bit stream is not free of redundancy, such that violations of syntactic or semantic constraints will usually occur quickly after a loss of synchronization [13], [30]–[32]. For example, the decoder might not find a matching VLC word in the code table (a syntax violation) or detect that the decoded motion vectors, DCT coefficients, or quantizer step sizes exceed their permissible range (semantic violations). Additionally, the accumulated run that is used to place DCT coefficients into an 8×8 block might exceed 64, or the number of MB's in a GOB might be too small or too large. Especially for severe errors, the detection of errors can be further supported by localizing visual artifacts that are unlikely to appear in natural video signals.

Recently, more advanced techniques for improved resynchronization have been developed in the context of MPEG-4. Among several error resilience tools, data partitioning has been shown to be effective [13]. Especially when combined with reversible VLC (RVLC), which allow bit streams to be decoded in either the forward or reverse direction, the number of symbols that have to be discarded can be reduced significantly. Because RVLC's can be matched well to the statistics of image and video data, only a small penalty in coding efficiency is incurred [33], [34]. A recently proposed technique can even approach the efficiency of Huffman codes by combining a prefix and suffix code word stream by delayed XORing [35]. Another elegant technique that is not part of any current video coding standard has been proposed by Kingsbury *et al.* as error-resilient entropy coding (EREC) [36]. Similar to

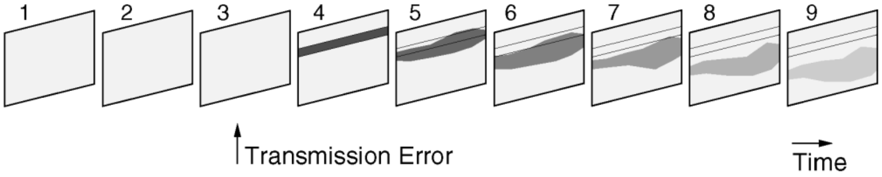


Fig. 2. Illustration of spatiotemporal error propagation.

data partitioning, a reordering of the bit stream is involved. Instead of clustering all symbols of the same type into one partition, EREC reorganizes VLC image blocks such that each block starts at a known position within the bit stream and the most important information is closest to the synchronization points.

However, even when GOB-headers are placed more frequently or more advanced techniques like data partitioning or EREC are applied, a certain amount of data has to be discarded when decoding an erroneous bit stream. The corresponding region in the image is then marked as lost and further processed by error concealment.

B. Error Concealment

The severeness of residual errors can be reduced if error concealment techniques are employed to hide visible distortion as well as possible. Since typically an entire GOB, or at least a significant part of it, is affected (i.e., 16 successive luminance lines) spatial interpolation is less efficient than temporal extrapolation. Only in the case of very complex motion or scene cuts, it can be advantageous to rely on the spatial correlation in the image [37], [38], or switch between temporal and spatial concealment [39], [32]. In the simplest and most common approach, previous frame concealment, the corrupted image content is replaced by corresponding pixels from the previous frame. This simple approach yields good results for sequences with little motion [30]. However, severe distortions are introduced for image regions containing heavy motion.

If data partitioning and strong error protection for the motion vector is used, one might rely on the transmitted motion vectors for motion-compensated concealment. If motion vectors are lost, they can be reconstructed by appropriate techniques, for example, by spatial interpolation of the motion vector field [40], which can be enhanced by additionally considering the smoothness of the concealed macroblock along the block boundaries [41], [42]. The interested reader is referred to [8] for a comprehensive overview of concealment techniques. All the feedback-based error control approaches discussed in the sequel benefit similarly from better concealment. Hence, it suffices to select one technique, and we present experimental results for the simple previous frame concealment in the following.

C. Interframe Error Propagation

Errors remaining after concealment propagate to successive frames and remain visible for a longer period of time, which makes the resulting artifacts particularly annoying. In addition, the accumulation of several errors can result in very poor image quality, even if the individual errors are small. Fig. 2 illustrates the typical transmission error effects for the loss of one GOB in frame 4. Not only does the error propagate temporally, but it also spreads spatially due to motion-compensated prediction. To some extent, the impairment caused by transmission errors decays over time due to leakage in the prediction loop. Leaky prediction is a well-known technique to increase the robustness of DPCM systems by attenuating the energy of the prediction signal [43]. For hybrid video coding, leakage is introduced by spatial filtering, as discussed in detail in the following.

Fig. 3 quantitatively illustrates the interframe error propagation after the loss of one GOB when previous frame concealment is used. The QCIF sequence *Foreman* is coded with H.263 at 100 kb/s and 12.5 f/s, resulting in an average peak signal-to-noise ratio (PSNR) of about 34 dB in the error-free case. Using the reconstructed frames at the encoder as the baseline, the loss ΔPSNR in signal-to-noise ratio is calculated for each reconstructed frame at the decoder output. The nine dotted curves correspond to each of the nine GOB's in the fifth encoded frame. Obviously, the error magnitude is highly image content dependent. The solid line represents the averaged result, indicating that an average residual loss of approximately 1 dB still remains in the sequence after 3 s due to interframe error propagation.

The decay of propagated errors is determined by two effects:

- some blocks are encoded in INTRA mode, i.e., without reference to the previous frame;
- repeated spatial filtering in the motion-compensated predictor, especially for half-pixel interpolation, attenuates the high spatial frequency components of the superimposed transmission error.

We can isolate the influence of these effects by experimental results shown in Fig. 4. We performed five simulations (A–E) with a baseline H.263 coder and decoder under different constraints with respect to spatial filtering and the use of INTRA mode. The general simulation conditions are identical to those in Fig. 3. All macroblocks are encoded in the INTER mode, except for simulation E, where nine out of 99 macroblocks per frame are randomly coded in INTRA mode. In simulation A, the subpixel fractions of the vector are forced to (0,0), i.e., motion compensation is carried out with integer–pixel accuracy and no spatial interpolation is required. In this case, the error due to a lost GOB does not decay over time. In simulation B, the subpixel fractions of the motion vector are forced

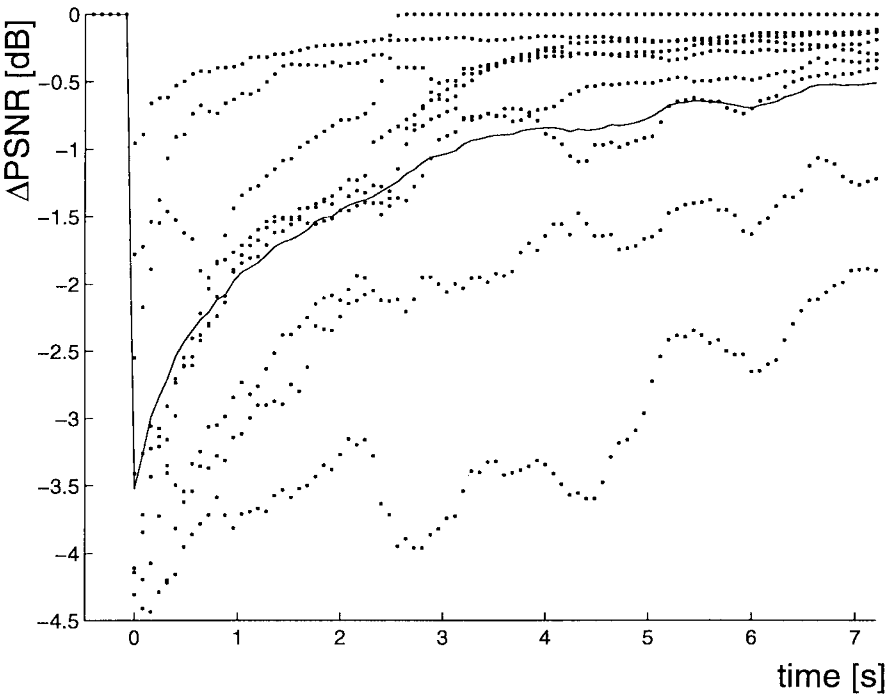


Fig. 3. Loss in SNR of the decoded video signal after previous frame concealment of one GOB.

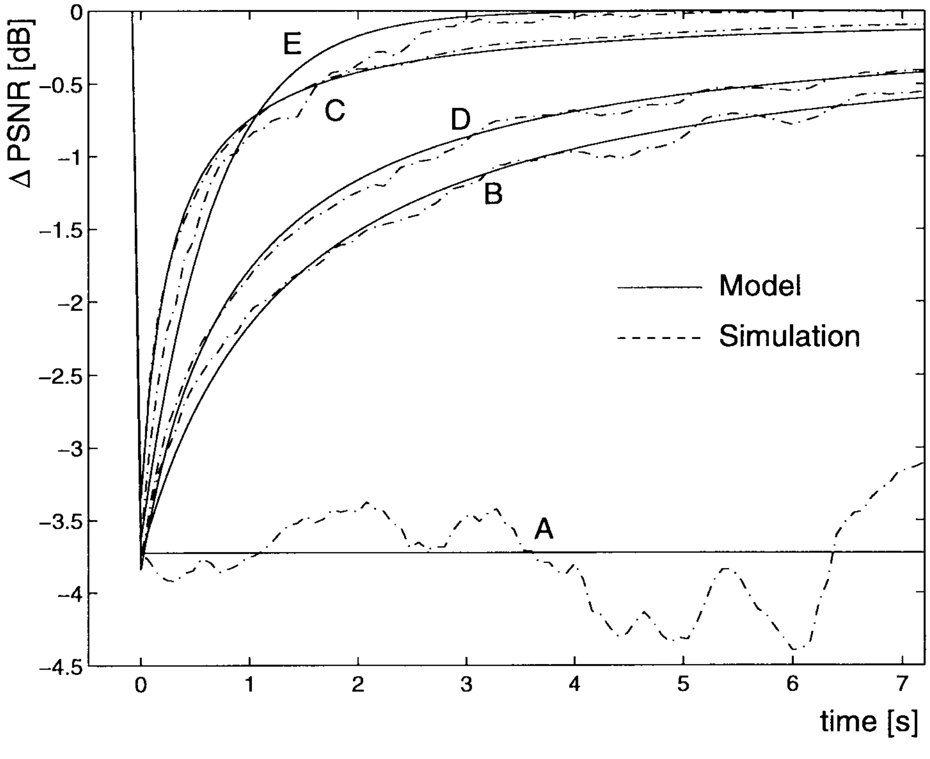


Fig. 4. Loss in PSNR of the decoded video-signal for different spatial interpolations (SI) in the prediction loop. A: no SI, B: horizontal SI, C: horizontal and vertical SI, D: SI not altered, E: SI not altered with 10% INTRA update.

to (0.5, 0.0), such that horizontal interpolation has to be performed for the entire image. Similarly, in simulation C the subpixel fractions are forced to (0.5, 0.5), such that horizontal and vertical interpolation must be performed for the entire image. H.263 employs bilinear interpolation for sub-pixel motion compensation, which acts as a low-pass filter, and the interframe error propagation decays faster, if more severe low-pass filtering is applied. For the remaining simulations D and E, the subpixel fractions are not altered but remain as selected by the encoder for best compression performance. Curve D is therefore a mixture of A, B, and C, while curve E decays faster then D due the use of intraframe coding.

The experimental curves shown in Fig. 4 can be approximated quite accurately by an analytical model that we derive in the Appendix. To match the model with experimental results, two parameters are necessary to describe the introduced error which remains after previous frame concealment. The first parameter is the average energy of the error, and the second parameter describes the shape of the power spectral density (PSD) when approximated by a Gaussian function. Both parameters can be derived from the true PSD of the error which has significant influence on interframe error propagation. The model predicts that for case B, the energy decreases with $\sqrt{1/t}$ for large t, or with $1/t$ for case C. Such analytical models are important to understand interframe error propagation and can be built into error tracking algorithms, as discussed in Section IV.

Spatial filtering in the motion-compensated predictor is a necessary ingredient for good compression performance of a hybrid coder [44], [45]. Even with integer-pixel accurate motion compensation, a "loop filter" should be employed. For example, in H.261, which uses integer-pixel motion compensation, the PSNR gain due to the loop filter is up to 2 dB [20]. While error recovery is also improved significantly at the same time, this is really a side-effect, and, as Fig. 4 shows, the leakage in the DPCM loop of standardized video codecs by itself is not strong enough for error robustness. For this purpose, additional leakage could be introduced at the cost of coding efficiency [46], [40]. On the other hand, quick error recovery is also possible when INTRA coding is used as illustrated by curve E.

A safe method to stop interframe error propagation that is used in MPEG is the regular insertion of I-frames, i.e., pictures that are encoded entirely in INTRA mode. Unfortunately, I-frames typically require several times more bits than P-frames (the MPEG term for pictures encoded with reference to previous frames). While this is acceptable for higher bit-rate applications, or even necessary for broadcasting, where many receivers need to resynchronize at random times, the use of the INTRA mode should be restricted as much as possible in point-to-point transmission at low bit rates, as typical for mobile multimedia networks. Feedback-based methods described in the following section can efficiently minimize the use of the INTRA mode and therefore maintain higher coding efficiency for hostile channels.

IV. ERROR MITIGATION BY FEEDBACK

As shown in the previous section, the remaining distortion after error concealment of corrupted image regions may remain visible in the image sequence for several seconds. In this section, we discuss error resilience techniques that utilize a feedback channel from the receiver to the transmitter. Such a feedback channel indicates which parts of the bit stream were received intact and/or which parts of the video signal could not be decoded and had to be concealed. Depending on the desired error behavior, negative acknowledgment (NACK) or positive acknowledgment (ACK) messages can be sent. Typically, an ACK or NACK refers to a series of macroblocks or an entire GOB. NACK's require a lower bit rate than ACK's, since they are only sent when errors actually occur, while ACK's have to be sent continuously. In either case, the requirements on the bit rate are very modest compared to the video bit rate of the forward channel. The feedback message is usually not part of the video syntax but transmitted in a different layer of the protocol stack where control information is exchanged. For example, in conjunction with H.263, ITU-T Recommendation H.245 [25] allows reporting of the temporal and spatial location of MB's that could not be decoded successfully and had to be concealed. Since the information is transmitted using a retransmission protocol, the error-free reception is guaranteed. However, additional delay my be introduced in the case of errors. In the following we assume that ACK's/NACK's are received error free after a relatively large round trip delay, e.g., 300 ms. This delay covers several retransmission attempts and may be considered as a worst case estimate for the actual delay. An alternative approach would be to sacrifice the reliable transmission of acknowledgment information at the advantage of reduced delay. Though this approach may be advantageous in some situations, we only consider reliable transmission of ACK's/NACK's in the following.

A. *Error Tracking*

The error tracking approach uses the INTRA mode for some MB's to stop interframe error propagation but limits its use to severely affected image regions only. During error-free transmission, the more effective INTER mode is used and the system therefore adapts effectively to varying channel conditions. This is accomplished by processing the NACK's from a feedback channel in the coding control of the encoder (Fig. 1). Based on the information of an NACK, the encoder can reconstruct the resulting error distribution in the current frame as described below. The coding control of a forward-adaptive encoder can then effectively stop interframe error propagation by selecting the INTRA mode whenever a MB is severely distorted. On the other hand, if error concealment was successful and the error of a certain MB is small, the encoder may decide that INTRA coding is not necessary. For severe errors however, a large number of MB's is selected, and the encoder may have to use a coarser quantizer to maintain a constant frame rate and bit rate. In this case, the overall picture quality decreases with a higher frequency of NACK's. Unlike retransmission techniques such as ARQ, error tracking does not increase the delay between encoder and decoder. It is therefore particularly suitable for applications that require a short latency.

Fig. 5 illustrates error tracking for the same example as in Fig. 2. As soon as the NACK is received with a system-dependent round-trip delay, the impaired MB's are determined and error propagation can be terminated by INTRA coding these MB's (frames 7–9). Fig. 6 shows the averaged PSNR loss for an assumed round-trip delay

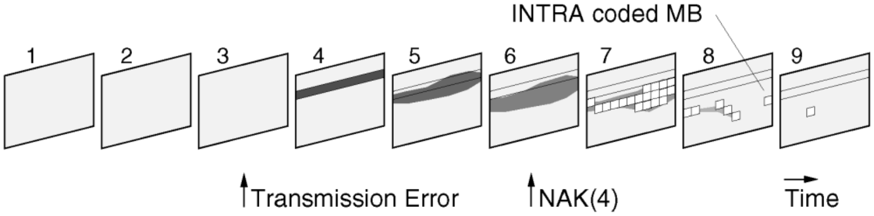


Fig. 5. Illustration of error propagation when error tracking is used.

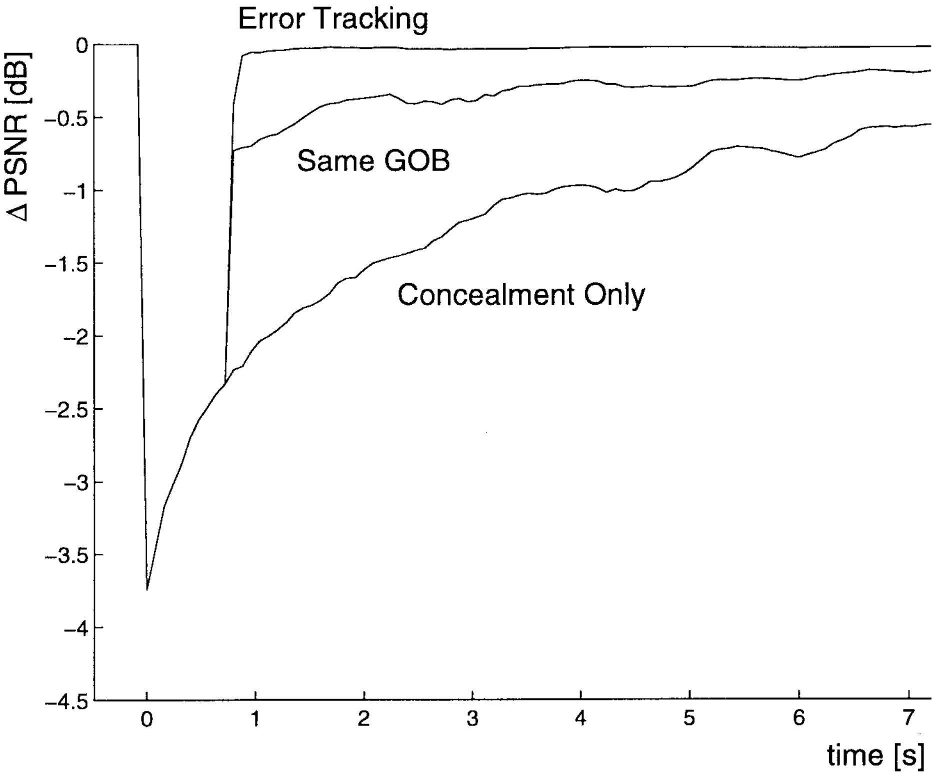


Fig. 6. Error recovery with feedback channel.

of 800 ms. The same simulation conditions as in Fig. 3 are used. Compared to the case without error tracking, the picture quality recovers rapidly as soon as INTRA coded MB's are inserted into the bit stream. A longer round-trip delay just results in a later start of the error recovery. Considering the slow recovery for concealment only, NACK's may still be useful after several seconds. In order to illustrate the importance of actually tracking the shifting location of the error, we also show results for the simple same GOB strategy, where the error is assumed to remain in the GOB were it originally occurred. When errors are dragged out of the original GOB along with vertical motion of picture contents, the same GOB strategy cannot remove it successfully, and annoying artifacts remain. Only when error tracking is employed, the propagation is stopped effectively. This is also demonstrated in Fig. 7, which shows example frames of the H.263 encoded test sequence *Foreman* directly after the NACK is received.

In order to reconstruct the interframe error propagation that has occurred at the decoder, the encoder could store its own output bit stream and decode it again, taking into account the reported loss of GOB's. While this approach is not feasible for a real-time implementation, it illustrates that the encoder, in principle, possesses all the information necessary to reconstruct the spatiotemporal error propagation at the decoder, once the NACK's have been received. For a practical system, the interframe error propagation has to be estimated with a low-complexity algorithm, as described in [47]–[49] for a macroblock-based coder, such as H.263.

The basic idea of the low-complexity algorithm is to carry out the error tracking with macroblock resolution rather than pixel resolution. This is sufficient since the INTRA/INTER mode decision at the coder and the error concealment decision at the decoder are carried out for entire MB's as well. In a cyclical buffer that covers all the MB's of the last several frames, the spatial overlap of MB's in successive frames due to motion-compensated prediction is stored, along with the error energy that would be introduced if concealment had to be used. If an NACK is received that indicates an error a few frames back, this

Fig. 7. Reconstructed frames of test sequence *Foreman*: (a) frame 90 after previous frame concealment of two GOB's in frame 75; (b) as (a) with INTRA update in frame 90 according to Same GOB strategy; (c) as (a) with INTRA update in frame 90 according to error tracking strategy; (d) frame 90 without GOB loss in frame 75.

error energy is "released" and "ripples" through the directed graph of frame-to-frame dependencies to the macroblocks of the current frame. The interframe error propagation model derived in the Appendix can be incorporated for more accurate prediction. Since all calculations are carried out on the MB level, the computational burden and memory requirements are small compared to the actual encoding of the video. For example, at QCIF resolution, there are only 99 MB's in each frame, as opposed to 38 016 luminance and chrominance samples.

The above error tracking scheme is a refinement of Wada's "selective recovery method" [50]. Wada's method marks all potentially damaged image blocks by one bit and prevents their use for interframe prediction. The method described here also calculates the severeness of the impairment and can thus use the extra bits required for INTRA coding more sparingly. Note that the coder has to know the decoder's concealment technique for that. On the other hand, Wada's method only considers the worst possible interframe error propagation and hence does not require an agreed concealment technique.

Error tracking is particularly attractive since it does not require any modifications of the bit stream syntax of the motion-compensated hybrid coder. It is therefore fully compatible with standards such as H.261, H.263, or MPEG. The ITU-T recommends using previous frame concealment and error tracking with baseline H.263 and has included an informative appendix with Recommendation H.263. In addition, minor extensions of the H.245 control standard were adopted to include the appropriate NACK messages.

B. Error Confinement

While interframe error propagation can be tracked reliably and with low complexity, a number of proposals for feedback-based error control technique do not rely on this technique but confine the error to a well-defined subregion of the frame instead. For example, in MPEG-4, arbitrarily shaped video object planes (VOP's) can be

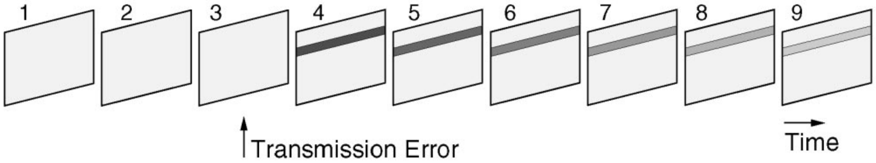


Fig. 8. Illustration of error propagation when error confinement is used.

encoded independently and superimposed at the decoder [13], [18]. If an error occurs in the bit stream of one VOP, the interframe error propagation is confined to that VOP since other VOP's will not refer to it for motion-compensated prediction. Basically, the video sequence is partitioned into independent subvideos.

Such a subvideo technique has also been included in H.263+ as an optional extension. The "independent segment decoding" (ISD) mode is described in Annex R of H.263. It can also be combined with slices, but we restrict the discussion to the case in which a segment is identical to a GOB. In the ISD mode, each GOB is encoded as an individual picture (or subvideo) independently from other GOB's. In particular, all GOB boundaries are treated just like picture boundaries. This approach significantly reduces the efficiency of motion compensation, particularly for vertical motion, since image content outside the current GOB must not be used for prediction. To reduce the loss of coding efficiency the ISD mode is therefore often combined with the "unrestricted motion vector" (UMV) mode which allows motion vectors pointing outside the coded picture area by extrapolating the image (or subvideo) borders. In spite of the UMV mode, typical losses in PSNR in the range from 0.2 to 1.0 dB often have to be accepted.

In case of transmission errors, the ISD mode assures that errors inside a GOB will not propagate to other GOB's, as illustrated in Fig. 8. Of course, the ISD mode alone does not solve the problem of temporal error propagation. It only simplifies keeping track of the error effects. The error propagation itself must be combatted by feedback based INTRA updates, or by the use of reference picture selection (RPS).

C. Reference Picture Selection

Rather than switching to INTRA mode at the encoder to stop interframe error propagation at the decoder, the coder could also encode the current frame with reference to a previous frame that has been successfully decoded. This RPS approach can lower the excess bit-rate due to NACK-induced INTRA coding [51], [52].

H.263+ has included RPS as an option, described in Annex N. As for the discussion of the ISD mode, we again consider the case that GOB headers are used. Then, in H.263, the reference picture is selected on a GOB basis, i.e., for all MB's within one GOB the same reference picture is used. In order to stop error propagation while maintaining the best coding efficiency, the last frame available without errors at the decoder should be selected. The RPS mode can be combined with the ISD mode to avoid spatial error propagation, or, for better coding efficiency, with an error tracking algorithm.

RPS can be operated in two different modes. When the encoder receives only negative acknowledgments, the operation of the encoder is not altered during error-free transmission, and the GOB's of the previous frame are used as a reference. After a transmission error, the decoder sends an NACK for an erroneous GOB and thereby requests that older, intact frames provide the reference-GOB. The typical transmission error effects are illustrated in Fig. 9, where the selection of reference-GOB's is indicated by arrows. Note that the use of the ISD mode is assumed and the indicated selection is only valid for the erroneous GOB. The encoder receives an NACK for frame 4 before the encoding of frame 7. The NACK includes the explicit request to use frame 3 for prediction, which is observed by the encoder. Similar to the error tracking approach, the quality degrades until the requested GOB arrives at the decoder, i.e., for the period of one round trip delay. Therefore, the loss of picture quality after a transmission error and the recovery after receiving an NACK behaves very similarly to basic error tracking. The advantage of the RPS mode versus simply switching to INTRA mode lies in the increased coding efficiency. Fewer bits are needed for encoding the motion-compensated prediction error than for the video signal itself, even if the time lag between the reference frame and the current frame is several frame intervals.

In the positive acknowledgment mode, all correctly received GOB's are acknowledged and the encoder uses only confirmed GOB's as a reference. Since the encoder has to use older reference pictures for motion-compensated prediction with increasing round-trip time, the coding performance decreases, even if no transmission errors occur. On the other hand, error propagation is avoided entirely since only error free pictures are used for prediction.

RPS requires additional frame buffers at the encoder and decoder to store enough previous frames to cover the maximum round trip delay of NACK's or ACK's. In the NACK mode, the storage requirements of the decoder can be reduced to two frame buffers. Furthermore, if only error-free GOB's shall be displayed, one frame buffer is sufficient. In the ACK mode no such storage reduction is possible, unless a combination of both modes is allowed [52]. Increased storage requirements may still be considered a problem for inexpensive mobile terminals for some time.

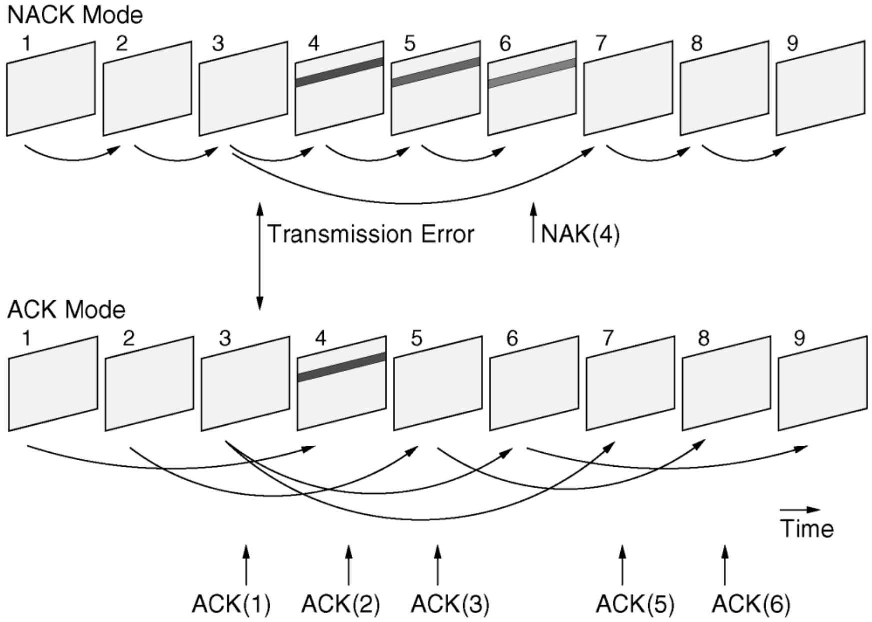


Fig. 9. Illustration of error propagation when RPS is used together with error confinement.

On the other hand, there are proposals for increased coding efficiency in low bit-rate video codecs which use several or even many previous frames for prediction [53]–[55]. When using RPS, the additional frames can then be used to simultaneously increase error robustness.

The different techniques for feedback-based error control that have been discussed in this section can be used in various combinations. Either error propagation is stopped by an INTRA update of affected regions or by selecting an error-free reference frame. Either the ACK mode or the NACK mode can be used for the latter option. Furthermore, either error tracking is used to identify affected regions or error propagation is confined. The advantages and disadvantages of possible combinations are summarized in Table 1.

V. VIDEO TRANSMISSION OVER A WIRELESS DECT CHANNEL

In this section, we illustrate the performance of feedback-based error mitigation in a mobile environment by simulating the transmission over a wireless channel. The channel model is related to the Digital European Cordless Telecommunications (DECT) standard, which provides a wide range of services for cordless personal communications [56], [57]. Our simulations are based on bit error sequences that are generated assuming Rayleigh fading and a Doppler frequency of 62 Hz. Carrier-to-noise ratios (E_b/N_0) ranging from 20 to 30 dB are considered with corresponding bit error rates as summarized in Table 2. The bit error sequences exhibit severe burst errors at a total bit rate of 80 kb/s, corresponding to the double slot format of DECT.

We apply FEC directly to the video bit stream using a BCH code of block size $n = 255$ bits with $k = 179$ information bits and $t = 10$ correctable errors per block. Due to the burstyness of the channel, not all errors can be corrected, and a significant block error rate remains, as also shown in Table 2. If the errors within a block cannot be corrected by FEC, all affected GOB's of the video bit stream are discarded and the video decoder invokes error concealment. We compare the error robustness of a baseline H.263 codec with and without error tracking to an H.263+ codec that uses NACK's, reference picture selection and error confinement by ISD in conjunction with the UMV mode. The feedback channel is assumed to be error-free with a constant delay of 100 ms. The round-trip delay, measured from encoding a frame to receiving an NACK, is about 300 ms due to the processing delay and buffering for constant transmission bit rate.

Twelve seconds of a typical videophone sequence *(Mother and Daughter)* are encoded at 12.5 f/s and transmitted over 30 different realizations of each test channel. Fig. 10 shows the average PSNR of the reconstructed frames at the encoder (after coding) and decoder (after transmission). The results are averaged values for all frames and realizations at each E_b/N_0. Note that for error tracking the picture quality after coding increases with increasing E_b/N_0 because fewer NACK's are received at better channel conditions and fewer MB's have to be coded in INTRA mode. Similarly, in the RPS mode, fewer GOB's have to be predicted with a higher time lag at better channel conditions. Both channel-adaptive approaches perform significantly better than the baseline mode of H.263 without feedback (NO). For lower E_b/N_0 the performance of the RPS mode in terms of $\overline{\text{PSNR}}$ is slightly superior to error tracking, since the INTRA mode can often be avoided. For higher E_b/N_0 this situation is reversed, because motion compensation is less efficient with the ISD mode. In summary, both channel-adaptive schemes perform very similarly and clearly outperform the nonadaptive scheme.

Table 1
Summary of Feedback-Based Error Control

	INTRA update	RPS NACK	RPS ACK
General			
Frames affected by interframe error propagation	N-1	N-1	0
Extra frame buffers at encoder/decoder	0/0	N/0 (N/1)	N/N
Error Tracking			
Coding efficiency w/o transmission errors	not affected	not affected	decreases with rtd
Loss of coding efficiency in case of transmission errors	moderate	small	very small
Computational complexity	low	low	low
Standards compliance	H.261, H.263, MPEG-1, MPEG-2, MPEG-4	H.263+ (Annex N)	H.263+ (Annex N)
Error Confinement			
Coding efficiency w/o transmission errors	somewhat reduced	somewhat reduced	somewhat reduced, decreases further with rtd
Loss of coding efficiency in case of transmission errors	significant	small	very small
Computational complexity	very low	very low	very low
Standards compliance	H.263+ (Annex R) MPEG-4	H.263+ (Annex N+R)	H.263+ (Annex N+R)

rtd: round-trip delay
N: number of encoded frames during one round-trip delay

Table 2
Summary of Test Channel Parameters

SNR [dB]	Bit Error Rate	Block Error Rate
20	0.002578	0.017075
22	0.001646	0.011250
24	0.001025	0.007575
26	0.000644	0.004675
28	0.000390	0.002775
30	0.000234	0.001725

VI. FEEDBACK-BASED ERROR CONTROL FOR PRECOMPRESSED VIDEO

In the above channel-adaptive schemes, acknowledgment information is incorporated during the encoding process. Therefore, these approaches are particularly relevant for conversational services or surveillance where video is encoded while being transmitted. However, feedback-based error mitigation can also be used for applications involving precompressed stored video, like video-on-demand for mobile clients. In this case, a direct influence on the coding control is no longer possible. To still adapt the bit stream to the error conditions of the channel, multiple bit streams have to be stored that are assembled dynamically for transmission [58], [59].

Assume that two bit streams are stored at the video server. The first stream (P-stream) results from regular encoding of the video sequence and consists of INTER-coded frames (P-frames). In normal operation, the video server transmits the P-stream only. The second stream (I-stream) consists of INTRA-coded frames (I-frames) and is used for resynchronization after transmission errors. The I-stream can be encoded at a lower frame rate to reduce the storage requirements at the video server. Upon receiving an NACK, the video server uses an error tracking algorithm to identify damaged image regions. These regions are then extracted from the next I-frame in the I-stream and inserted into the transmitted bit stream at the corresponding position. If predictive encoding of symbols (e.g. motion vectors) is used, as is the case for H.263 or MPEG, the bit stream has to be assembled on the slice or GOB level rather than the MB level. In this fashion, error propagation is limited almost as for real-time encoding. Of course, the selection of INTRA-coded regions happens at a coarser level.

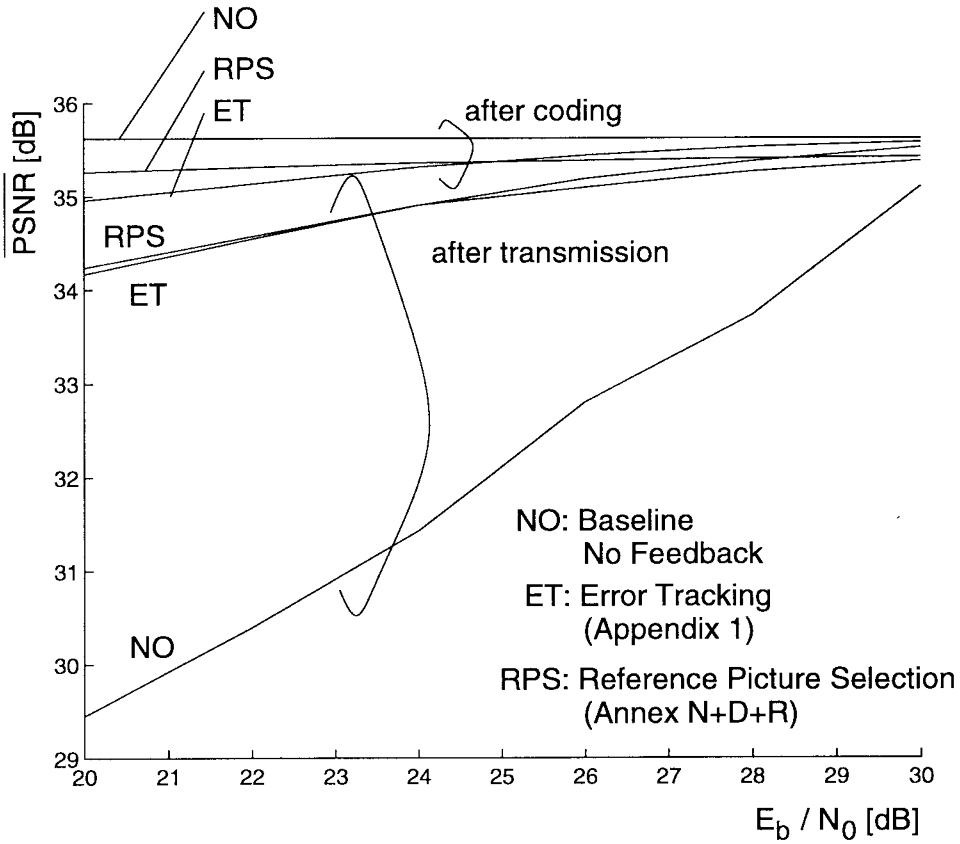


Fig. 10. Average PSNR of reconstructed frames at encoder (after coding) and decoder (after transmission) for the simulated transmission of H.263 coded video over a wireless DECT channel.

Note that the NACK-induced switching between the I-stream and the P-stream stops error propagation, but introduces a new mismatch error which results from the difference of the reconstructed P-frame and its corresponding reconstructed I-frame. If the difference is only minor, it will fade away over the following few P-frames as discussed in Section III-C. However, when there is a considerable difference and only weak filtering is applied in the coding loop, the mismatch error will cause annoying visual artifacts. To reduce the mismatch error, the reconstructed I-frame should be as close as possible to its corresponding P-frame. To achieve this, the I-stream is encoded using the reconstructed frames of the P-stream as the input instead of the original sequence.

Again, we demonstrate the feasibility of the approach in the framework of H.263. Fig. 11 shows the average loss of PSNR due to a burst error with the familiar recovery process in the case of no available feedback information. As soon as the NACK arrives and INTRA-coded GOB's are inserted, the picture quality improves similarly to the simulation results shown in Fig. 6. If the I-stream is generated by encoding the original sequence (case A), a good picture quality can be observed directly after the insertion. However, after several frames from the P-stream are transmitted, the mismatch between the reconstructed I-frame and the corresponding reconstructed P-frame dominates. This effect can be reduced if the I-stream is generated by encoding the reconstructed frames from the P-stream (case B).

Interestingly, this approach also offers a very efficient solution to random access. In the common approach, I-frames are inserted periodically into the bit stream to provide random access. In this case, I-frames also have to be transmitted during normal-speed playback, even though they do not provide any additional functionality. However, when the bit stream is assembled dynamically at the server, the higher efficiency of the INTER mode can be fully exploited during normal-speed playback. Random access, fast forward, and fast reverse can still be supported, because I-frames may be inserted whenever requested by the client. Note that this approach is fully standard compliant with H.261, H.263, or MPEG and requires only little additional complexity and storage overhead at the video server. It has also been extended to support bit rate and resolution scalability [60].

VII. CONCLUDING REMARKS

In this paper, we have reviewed feedback-based techniques for robust transmission of video in mobile multimedia systems. We argue that transcoders significantly increase complexity and add unacceptable delay, and that therefore end-to-end error control has to be used. Each decoder therefore has to make provisions for error detection, resynchronization and error concealment. Additionally, we

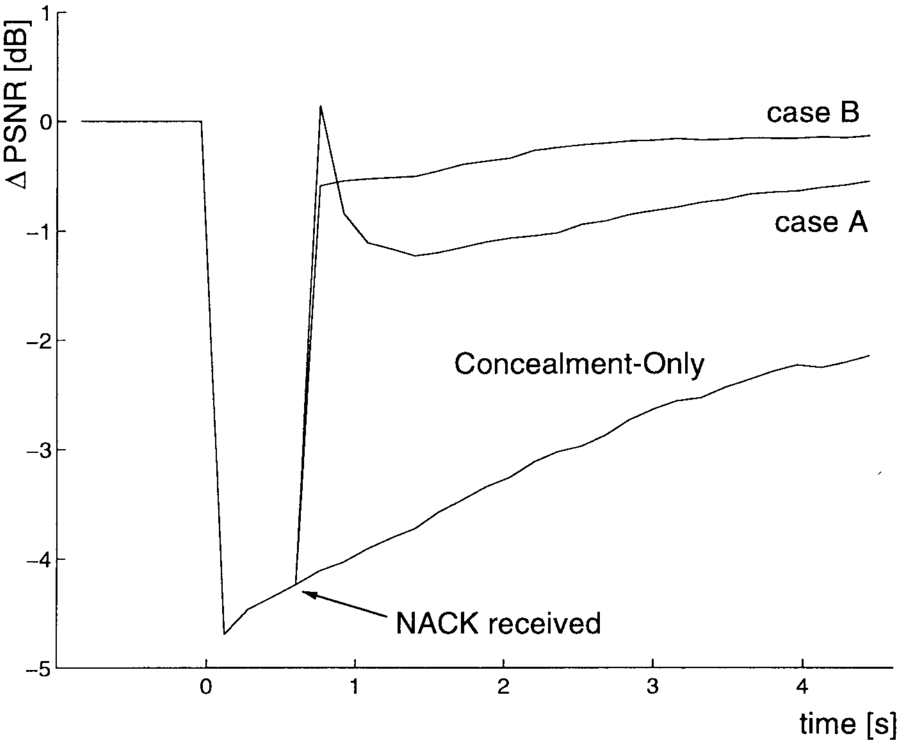


Fig. 11. Loss in PSNR of decoded video signal for inserting I-stream information into a P-stream in a video server application. Case A: I-stream encoded from original sequence. Case B: I-stream encoded from reconstructed P-stream.

advocate the intelligent processing of acknowledgment information by the coding control to adapt the source coder to the channel.

Error tracking allows the encoder to estimate interframe error propagation accurately and adapt its encoding strategy to mitigate the effects of past transmission errors. While error tracking is possible with low complexity, it can be simplified by error confinement techniques, at the expense of compression efficiency. Compression efficiency in case of transmission errors can be improved, however, if multiple previous pictures are kept in an RPS scheme. The simulation of a video transmission over a wireless channel indicates that different channel-adaptive schemes perform very similar but have a significant advantage over schemes that are not feedback-based.

The ITU-T Study Group 16 has adopted feedback-based error control in their effort toward mobile extensions of the successful Recommendation H.263. The first version of H.263 included basic error tracking as described in Section IV-A. The second version, which is informally known as H.263+, was adopted by ITU-T in February 1998. Among many other enhancements, it contains two new options supporting RPS (Annex N) and ISD (Annex R), both discussed in more detail in this article. Future enhancements, e.g., data partitioning, unequal error protection, and reversible variable length coding, are under consideration for future versions of the standard, informally known as H.263++ and H.26L.

Feedback schemes are suitable for interactive, individual communications, but they have inherent limitations outside this domain. They are particularly powerful if the round-trip delay is short. If the round-trip delay increases they become less efficient and ultimately useless. Also, feedback schemes are particularly advantageous for point-to-point communications. For schemes like error tracking that use INTRA updates to stop error propagation, extensions to few users are possible at a loss in performance. In the case of RPS, multipoint communication may even become impossible, because different decoders may require different reference frames. On the other hand, the objection that feedback-based source coding can only be used with real-time encoding is not valid, as shown in Section VI.

While most of the discussed feedback-based error control schemes are pragmatic engineering solutions to a problem at hand, there is an unsolved theoretical problem of fundamental importance underneath. Traditionally, error control has been performed without regard to the contents of the bit stream, e.g., by FEC or transport layer retransmission schemes. Using channel-adaptive source coding for robust transmission is a relatively new idea. In our work, we have learned that well-designed mobile video systems should combine all these techniques, as exemplified in Section V. The tradeoffs in such a system design, however, are not at all well understood, and a general theoretical framework for feedback-based error control is needed.

APPENDIX
ANALYSIS OF INTERFRAME ERROR PROPAGATION

In the following we derive an analytical model for the distortion that is caused by transmission errors and propagates due to the recursive DPCM structure in the decoder. We are interested in the signal $v[x, y, t]$, which is the frame difference of the reconstructed frames at the encoder and decoder at time step t. We assume that the residual error $u[x, y]$ is introduced at $t = 0$ (after resynchronization and error concealment) such that $v[x, y, 0] = u[x, y]$. In particular, we are interested in the variance $\sigma_v^2[t]$ of the propagated error signal.

A. Two-Dimensional Case

First we consider the two-dimensional case with the discrete spatial and temporal variables x and t. When the decoder is regarded as a linear system $H_t(\omega_x)$ with parameter t, the variance of $v[x, t]$ can be obtained as

$$\sigma_v^2[t] = \frac{1}{2\pi} \int_{-\pi}^{+\pi} |H_t(\omega_x)|^2 \Phi_{uu}(\omega_x)\, d\omega_x \tag{1}$$

where Φ_{uu} is the power spectral density (PSD) of the signal $u[x]$. We assume that a spatial filter $F(\omega_x)$ is applied in each time step. Then the impulse response of the decoder $h_t[x]$ can be defined recursively as $h_t[x] = h_{t-1}[x] * f[x]$, where "$*$" denotes discrete spatial convolution and $f[x]$ is the impulse response of the filter. Based on the central limit theorem we expect $h_t[x]$ to be Gaussian for large t. Therefore, the magnitude of the transfer function of the decoder can be approximated in the base band $|\omega_x| < \pi$ by

$$|\hat{H}_t(\omega_x)| = \exp\left(-\frac{\omega_x^2 t \sigma_f^2}{2}\right) \tag{2}$$

where σ_f^2 is defined as

$$\sigma_f^2 = \sum_x x^2 f[x] - \left(\sum_x x f[x]\right)^2. \tag{3}$$

In the case of linear interpolation, as commonly used for half-pel motion compensation, we obtain $\sigma_f^2 = 1/4$. In addition to the Gaussian approximation for $H_t(\omega_x)$ we also approximate the PSD of the introduced error signal $u[x]$ by

$$\hat{\Phi}_{uu}(\omega_x) = \sigma_u^2 \sqrt{4\pi\sigma_g^2} \exp\left(-\omega_x^2 \sigma_g^2\right) \tag{4}$$

i.e., a Gaussian PSD with the energy σ_u^2. The parameter σ_g^2 determines the shape of the PSD and can be used to match (4) with the true PSD. With the given approximations for $|H_t(\omega_x)|$ and $\Phi_{uu}(\omega_x)$ we can solve (1) directly yielding

$$\hat{\sigma}_v^2[t] = \sigma_u^2 \sqrt{\frac{\sigma_g^2}{\sigma_g^2 + t\sigma_f^2}} = \sigma_u^2 \alpha[t] \tag{5}$$

where $\alpha[t]$ is the power transfer factor after t time steps.

B. Extension to Three-Dimensional Case

In the following we focus on spatial filtering caused by motion compensated prediction with half-pel accuracy and bilinear interpolation. For the extension of the above results to the three-dimensional case with the discrete variables x, y, and t, it has to be considered that individual image regions undergo different filtering operations. For each macroblock a different motion vector is selected, and depending on its sub-pel fractions either no filtering, only horizontal, only vertical, or horizontal and vertical spatial filtering is applied.

In the process of encoding, different combinations of filters are applied to an individual image region after t time steps. The probability that an image region is filtered h times horizontally and v times vertically can be described by a two dimensional probability density function (pdf) $p_t[h, v]$ with parameter t. For $t = 0$ no filter operations have yet been performed, i.e., $p_0[0, 0] = 1$ and otherwise zero. For $t = 1$ the pdf is given by $p_1[0, 0]$, $p_1[1, 0]$, $p_1[0, 1]$, and $p_1[1, 1]$, with all other probabilities being equal to zero. In general, these four probabilities are equal to $1/4$ in moving areas unless specific sub-pel fractions are enforced. For $t > 1$ the pdf can be defined recursively according to

$$p_t[h, v] = p_{t-1}[h, v] * p_1[h, v] \tag{6}$$

under the assumption that the subpel fractions are independent in each frame at a given location. With this assumption and the definition of the power transfer factor $\alpha[\cdot]$ in (5), we obtain the variance of the signal v for the three-dimensional case as

$$\hat{\sigma}_v^2[t] = \sigma_u^2 \sum_h \sum_v \alpha[h]\alpha[v] p_t[h, v]. \tag{7}$$

This equation does not consider INTRA coded macroblocks, which will cause a faster decrease in error energy. If the INTRA mode is selected randomly for n out of N macroblocks per frame, the effect on the variance can be modeled as an additional leakage $\beta = 1 - n/N$ resulting in

$$\hat{\sigma}_v^2[t] = \beta^t \sigma_u^2 \sum_h \sum_v \alpha[h]\alpha[v] p_t[h, v]. \tag{8}$$

REFERENCES

[1] E. Berruto, M. Gudmundson, R. Menolascino, W. Mohr, and M. Pizarroso, "Research activities on UMTS radio interface, network architectures, and planning," *IEEE Commun. Mag.*, vol. 36, pp. 82–95, Feb. 1998.

[2] D. Grillo, Ed., "Special section on third-generation mobile systems in Europe," *IEEE Personal Commun. Mag.*, vol. 5, pp. 5–38, Apr. 1998.

[3] P. Bahl and B. Girod, Eds., "Special section on wireless video," *IEEE Commun. Mag.*, vol. 36, pp. 92–151, June 1998.

[4] S. Lin, D. J. Costello, and M. J. Miller, "Automatic repeat error control schemes," *IEEE Commun. Mag.*, vol. 22, pp. 5–17, Dec. 1984.

[5] A. Heron and N. MacDonald, "Video transmission over a radio link using H.261 and DECT," in *Inst. Elect. Eng. Conf. Publications*, 1992, no. 354, pp. 621–624.

[6] M. Khansari, A. Jalali, E. Dubois, and P. Mermelstein, "Low bit-rate video transmission over fading channels for wireless microcellular systems," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 1–11, Feb. 1996.

[7] P. Cherriman and L. Hanzo, "Programable H.263-based video transceivers for interference-limited environments," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, pp. 275–286, June 1998.

[8] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: A review," *Proc. IEEE*, vol. 86, pp. 974–997, May 1998.

[9] R. Fischer, P. Mangold, R. M. Pelz, and G. Nitsche, "Combined source and channel coding for very low bitrate mobile visual

communication systems," in *Proc. Int. Picture Coding Symp. (PCS)*, Melbourne, Australia, Mar. 1996, pp. 231–236.

[10] K. Illgner and D. Lappe, "Mobile multimedia communications in a universal telecommunications network," in *Proc. SPIE Conf. Visual Communications and Image Processing (VCIP)*, vol. 2501, Taipei, Taiwan, May 1995, pp. 1034–1043.

[11] B. Girod, U. Horn, and B. Belzer, "Scalable video coding with multiscale motion compensation and unequal error protection," in *Multimedia Communications and Video Coding*, Y. Wang, S. Panwar, S. P. Kim, and H. L. Bertoni, Eds. New York: Plenum, 1996, pp. 475–482.

[12] R. M. Pelz, "An unequal error protected $p \times 8$ kbit/s video transmission for dect," in *Proc. IEEE Vehicular Technology Conf. (VTC)*, Stockholm, Sweden, June 1994, pp. 1020–1024.

[13] R. Talluri, "Error-resilient video coding in the MPEG-4 standard," *IEEE Commun. Mag.*, vol. 36, pp. 112–119, June 1998.

[14] *Video Coding for Low Bitrate Communication*, ITU-T Recommendation H.263, 1996.

[15] A. M. Tekalp, *Digital Video Processing*. Upper Saddle River, NJ: Prentice Hall, 1995.

[16] *Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5 Mbit/s (MPEG-1)*, International Standard ISO/IEC 11172-2, 1994.

[17] *Generic Coding of Moving Pictures and Associated Audio, Recommendation H.262 (MPEG-2)*, International Standard ISO/IEC 13818-2, 1995.

[18] *Coding of Audio-Visual Objects: Visual (MPEG-4 Video)*, Committee Draft ISO/IEC 14496-2, Oct. 1997.

[19] *Video Coding for Audiovisual Services at $p \times 64$ kbit/s*, ITU-T Recommendation H.261, 1990.

[20] B. Girod, N. Färber, and E. Steinbach, "Performance of the H.263 video compression standard," *J. VLSI Signal Processing: Syst. Signal, Image, Video Technol.*, vol. 17, pp. 101–111, Nov. 1997.

[21] *Terminal for Low Bitrate Multimedia Communication*, ITU-T Recommendation H.324, 1995.

[22] D. Lindberg, "The H.324 multimedia communication standard," *IEEE Commun. Mag.*, vol. 34, pp. 46–51, Dec. 1996.

[23] *Multiplexing Protocol for Low Bitrate Multimedia Communication*, ITU-T Recommendation H.223, 1996.

[24] N. Färber, B. Girod, and J. Villasenor, "Extensions of ITU-T recommendation H.324 for error-resilient video transmission," *IEEE Commun. Mag.*, vol. 36, pp. 120–128, June 1998.

[25] *Control Protocol for Multimedia Communication*, ITU-T Recommendation H.245, 1996.

[26] *Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s*, ITU-T Recommendation G.723.1, 1996.

[27] T. J. Ferguson and J. H. Rabinowitz, "Self-synchronizing Huffman codes," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 687–693, July 1984.

[28] W.-M. Lam and A. R. Reibman, "Self-synchronizing variable length codes for image transmission," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'92)*, vol. 3, San Francisco, CA, Mar. 1992, pp. 477–480.

[29] J. Hagenauer and P. Höher, "A Viterbi algorithm with soft-decision output and its applications," in *Proc. IEEE Global Telecommunications Conf. (GLOBECOM)*, Dallas, TX, Nov. 1989, pp. 47.1.1–47.1.7.

[30] C. Chen, "Error detection and concealment with an unsupervised MPEG2 video decoder," *J. Visual Commun. Image Representation*, vol. 6, no. 3, pp. 265–278, Sept. 1995.

[31] W.-M. Lam and A. R. Reibman, "An error concealment algorithm for images subject to channel errors," *IEEE Trans. Image Processing*, vol. 4, pp. 533–542, May 1995.

[32] J. W. Park, J. W. Kim, and S. U. Lee, "DCT coefficients recovery-based error concealment technique and its application to the MPEG-2 bit stream," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 845–854, Dec. 1997.

[33] J. Wen and J. D. Villasenor, "A class of reversible variable length codes for robust image and video coding," in *Proc. 1997 IEEE Int. Conf. Image Processing (ICIP)*, vol. 2, Santa Barbara, CA, Oct. 1997, pp. 65–68.

[34] J. Wen and J. D. Villasenor, "Reversible variable length codes for efficient and robust image and video coding," in *Proc. IEEE Data Compression Conf. (DCC)*, Mar. 1998, Snowbird, UT, pp. 471–480.

[35] B. Girod, "Bidirectionally decodable streams of prefix code words," *IEEE Commun. Lett.*, to be published.

[36] D. W. Redmill and N. G. Kingsbury, "The EREC: An error-resilient technique for coding variable-length blocks of data," *IEEE Trans. Image Processing*, vol. 5, pp. 565–574, Apr. 1996.

[37] Y. Wang, Q.-F. Zhu, and L. Shaw, "Maximally smooth image recovery in transform coding," *IEEE Trans. Commun.*, vol. 41, pp. 1544–1551, Oct. 1993.

[38] H. Sun and W. Kwok, "Concealment of damaged block transform coded images using projection onto convex sets," *IEEE Trans. Image Processing*, vol. 4, pp. 470–477, Apr. 1995.

[39] H. Sun and J. Zedepski, "Adaptive error concealment algorithm for MPEG compressed video," in *Proc. SPIE Conf. Visual Communications and Image Processing (VCIP)*, vol. 1818, Boston, MA, Nov. 1992, pp. 814–824.

[40] P. Haskell and D. Messerschmitt, "Resynchronization of motion compensated video affected by ATM cell loss," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, San Francisco, CA, Mar. 1992, pp. 545–548.

[41] W.-M. Lam, A. R. Reibman, and B. Lin, "Recovery of lost or erroneously received motion vectors," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal processing (ICASSP)*, vol. 5, Minneapolis, MN, Apr. 1993, pp. 417–420.

[42] P. Salama, N. Shroff, and E. J. Delp, "A fast suboptimal approach to error concealment in encoded video streams," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, vol. 2, Santa Barbara, CA, Oct. 1997, pp. 101–104.

[43] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, NJ: Prentice Hall, 1984.

[44] B. Girod, "The efficiency of motion-compensating prediction for hybrid coding of video sequences," *IEEE J. Select. Areas Commun.*, vol. SAC-5, pp. 1140–1154, Aug. 1987.

[45] ——, "Motion-compensating prediction with fractional-pel accuracy," *IEEE Trans. Commun.*, vol. 41, pp. 604–612, Apr. 1993.

[46] D. J. Connor, "Techniques for reducing the visibility of transmission errors in digitally encoded video signals," *IEEE Trans. Commun.*, vol. COM-21, pp. 695–706, June 1973.

[47] N. Färber, E. Steinbach, and B. Girod, "Robust H.263 compatible video transmission over wireless channels," in *Proc. Int. Picture Coding Symp. (PCS)*, Melbourne, Australia, Mar. 1996, pp. 575–578.

[48] E. Steinbach, N. Färber, and B. Girod, "Standard compatible extension of H.263 for robust video transmission in mobile environments," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 872–881, Dec. 1997.

[49] B. Girod, N. Färber, and E. Steinbach, "Error-resilient coding for H.263," in *Insights into Mobile Multimedia Communication*, D. Bull, N. Canagarajah, and A. Nix, Eds. New York: Academic, 1999, pp. 445–449.

[50] M. Wada, "Selective recovery of video packet loss using error concealment," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 807–814, June 1989.

[51] S. Fukunaga, T. Nakai, and H. Inoue, "Error resilient video coding by dynamic replacing of reference pictures," in *Proc. IEEE Global Telecommunications Conf. (GLOBECOM)*, London, U.K., Nov. 1996, vol. 3, pp. 1503–1508.

[52] Y. Tomita, T. Kimura, and T. Ichikawa, "Error resilient modified inter-frame coding system for limited reference picture memories," in *Proc. Int. Picture Coding Symp. (PCS)*, Berlin, Germany, Sept. 1997, pp. 743–748.

[53] T. Wiegand, X. Zhang, and B. Girod, "Motion-compensating long-term memory prediction," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Santa Barbara, CA, Oct. 1997, pp. 53–56.

[54] T. Wiegand, X. Zhang, and B. Girod, "Long-term memory motion-compensated prediction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, pp. 70–84, Feb. 1999.

[55] B. Girod, T. Wiegand, E. Steinbach, M. Flierl, and X. Zhang, "High-order motion compensation for low bit-rate video," in *Proc. Europ. Signal Processing Conf.*, Rhodes, Greece, Sept. 1998, pp. 253–256.

[56] J. E. Padgett, C. Günther, and T. Hattori, "Overview of wireless personal communications," *IEEE Commun. Mag.*, vol. 33, pp. 28–41, Jan. 1995.

[57] P. Wong and D. Britland, *Mobile Data Communications Systems*. Norwood, MA: Artech House, 1995.

[58] N. Färber, E. Steinbach, and B. Girod, "Robust H.263 compatible transmission for mobile video server access," in *Proc. 1st Int. Workshop Wireless Image/Video Communications*, Lough-

borough, U.K., Sept. 1996, pp. 8–13.

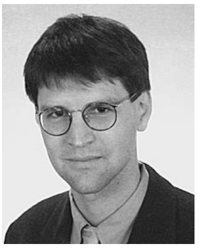
[59] N. Färber and B. Girod, "Robust H.263 compatible video transmission for mobile access to video servers," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, vol. 2, Santa Barbara, CA, Oct. 1997, pp. 73–76.

[60] B. Girod, N. Färber, and K. Stuhlmüller, "Internet video streaming," in *Multimedia Communications*, F. DeNatale and S. Pupolin, Eds. Berlin, Germany: Springer, 1999, pp. 547–556.

Bernd Girod (Fellow, IEEE) received the M.S. degree in electrical engineering from Georgia Institute of Technology, Atlanta, in 1980 and the Doctoral degree (with highest honors) from the University of Hannover, Germany, in 1987.

He is currently a Chaired Professor of Telecommunications in the Electrical Engineering Department and Head of the Telecommunications Institute I at the University of Erlangen-Nuremberg, Erlangen, Germany. Until 1987, he was a Member of the Research Staff at the Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, University of Hannover, working on moving image coding, human visual perception, and information theory. In 1988, he joined Massachusetts Institute of Technology, Cambridge, first as a Visiting Scientist with the Research Laboratory of Electronics, then as an Assistant Professor of Media Technology at the Media Laboratory. From 1990 to 1993, he was Professor of Computer Graphics and Technical Director of the Academy of Media Arts in Cologne, Germany, jointly appointed with the Computer Science Section of Cologne University. He was a Visiting Adjunct Professor with the Digital Signal Processing Group at Georgia Institute of Technology in 1993. Since 1997, he has codirected the newly found Telecommunications Laboratory at the University of Erlangen-Nuremberg. He served as the Chairman of the Electrical Engineering Department from 1995 to 1997, and as Director of the Center of Excellence "3-D Image Analysis and Synthesis" since 1995. He was a Visiting Professor with the Information Systems Laboratory of Stanford University, Stanford, CA, in 1997–1998 academic year. His research interests include image communication, video signal compression, human and machine vision, computer graphics and animation, as well as interactive media. For several years, he has served as a consultant to government agencies and companies, with special emphasis on start-up ventures. He has been a founder and Chief Scientist of Vivo Software, Inc., Waltham, MA (1993–1998), Chief Scientist of RealNetworks, Inc., Seattle, WA, (since 1998), and a board member of 8×8, Inc., Santa Clara, CA since 1996.

Dr. Girod was recently elected Fellow of the IEEE "for his contributions to the theory and practice of video communications."



Niko Färber (Member, IEEE) received the Dipl.-Ing. degree in electrical engineering from the Technical University of Munich, Germany, in 1993.

He was then with the Research Laboratory Mannesmann Pilotentwicklung, where he developed system components for satellite based vehicular navigation. In October 1993, he joined the Telecommunications Laboratory at the University of Erlangen-Nuremberg, Erlangen, Germany, where he is now a member of the Image Communication Group. He started his research on robust video transmission as a Visiting Scientist at the Image Processing Laboratory of University of California, Los Angeles, CA. Since then he has published several conference and journal papers on the subject and has contributed successfully to the ITU-T Study Group 16 efforts for robust extensions of the H.263 standard. He has also served as the Publicity Vice Chair for ICASSP-97 in Munich, Germany, and performed research for 8×8, Inc., Santa Clara, CA, and RealNetworks, Inc., Seattle, WA. In cooperation with RealNetworks, he holds a U.S. patent in the area of scalable video coding. His other research interests include combined source and channel coding and software only video coding.