# Standard-Compliant Enhancements of JVT Coded Video for Transmission over Fixed and Wireless IP

Thomas Stockhammer

Institute for Communications Engineering (LNT)
Munich University of Technology (TUM)
80290 Munich, Germany

Stephan Wenger

Communication and Operating Systems
Technical University Berlin
10587 Berlin, Germany

Ph.: + 49 89 28923474, Email: stockhammer@ei.tum.de, Web: www.lnt.ei.tum.de

**ABSTRACT**

*This paper describes standard compliant enhancements of the currently developed JVT coding algorithms for transmission over fixed and wireless IP-based networks. This includes a description of JVT coding algorithm and the adaptation to IP-based networks. The error resilience features within the JVT coding algorithm are presented in greater detail. Standard compliant encoder and decoder enhancements as well as the exploitation of network information to improve the quality in packet lossy environments are presented. Pointers and references to well-known and successfully applied error resilience schemes in prior video coding standards are provided. Different schemes are compared and appropriate experimental results based on common test conditions are presented and discussed.*

## 1 INTRODUCTION

Since 1997, the ITU-T's Video Coding Experts Group (VCEG) has been working on a new video coding standard with the internal denomination H.26L. In late 2001, the Moving Picture Expert Group (MPEG) and VCEG decided to work together as a Joint Video Team (JVT), and to create a single technical design for a forthcoming ITU-T Recommendation and for a new part of the MPEG-4 standard based on the current committee draft version 1 (CD-1) of JVT coding [2][1]. Since the meeting in May 2002 the technical specification is almost frozen and the presented concepts in this work will very likely be part of the final standard. The primary goals of the JVT project are *improved coding efficiency*, *improved network adaptation* and *simple syntax specification*. The syntax of JVT coding should permit an average reduction in bit rate by 50% compared to all previous standards for a similar degree of encoder optimization. Recent results show that this performance is easily achieved [3], [4]. This makes JVT coding an attractive candidate for many applications including fixed and wireless video transmission over the Internet Protocol (IP) [5]. However, to allow transmission in different environments not only coding efficiency is relevant, but also seamless and easy integration of the coded video into all current and possible future protocol and multiplex architectures and enhanced error resilience features are of major importance.

Previous video coding standards, such as MPEG-2 [5], MPEG-4 [7] and H.263 [8], were mainly designed for special applications or transport protocols usually in a circuit-switched or bit-stream oriented environment although they have been adapted to different transport protocols later on. JVT experts have taken into account transmission over packet-based networks in the video codec design from the very beginning. This is as IP-based standard compliant video transmission has attained significant interest recently. Typical applications include conversational services, such as videotelephony, and videoconferencing, streaming services or multimedia messaging services (MMS). In addition to traditional fixed Internet video conversational services, especially the video transmission over third generation mobile systems will be mainly packet-based. A mobile video codec design must minimize terminal complexity while remaining consistent with the efficiency and robustness goals of the design. For sufficient quality hardware support is necessary and this makes standard solutions in wireless environments even for streaming and MMS very attractive.

Video over IP is usually transported either by downloading complete bit streams (MMS) using reliable end-to-end protocols such as Transmission Control Protocol over IP (TCP/IP) [9], or by real-time transmission. The latter one applied for conversational or streaming services over IP networks usually employs IP [5] on the network layer, User Datagram protocol (UDP) [10] on the transport layer, and real-time transport protocol (RTP) [11] and accompanying RTP payload specifications, e.g. for MPEG-2 [12] for MPEG-4 [13] and for H.263 [14] on the application layer. However, UDP offers only a simple, unreliable datagram transport service: packets may get lost, duplicated, or re-ordered on their way from the source to the destination due network congestion, buffer overflows in intermediate routers or frame losses on mobile links.

The highly complex temporal and spatial prediction mechanisms included in modern video codecs like JVT coding result in catastrophic error propagation in case of packet losses. Then, the use of error resilience techniques in the source codec becomes important. Many schemes have been presented, investigated and assessed. For details, we refer to [15], [16], [17], [18], [19], [20] and references therein. The prime goal of his work is the adaptation of well-known and successfully applied techniques to JVT coding. Therefore, the remainder of this work is structured as follows: We will provide a concise overview over the JVT standard in Section 2. Section 3 presents error resilience tools in JVT coding. Section 4 and 5 discuss standard compliant enhancements of the JVT codec for increased packet loss resilience. Some concluding remarks will be provided in Section 6.

## 2 THE JVT VIDEO CODING STANDARD

### 2.1 Overview: JVT in Transport Environment

According to Figure 1, the JVT codec design distinguishes between two different conceptual layers, the Video Coding Layer (VCL), and the Network Abstraction Layer (NAL). Both the VCL and the NAL are part of the JVT standard. Additionally, interface specifications are required to different transport protocols, to be specified by the responsible standardization bodies. Furthermore, the exact transport and encapsulation of JVT data for different transport systems, such as H.320 [21], MPEG-2 Systems [5], and RTP/IP, are also

[1] All referenced standard documents can be accessed via anonymous ftp at ftp://standard.pictel.com/video_site, ftp://ftp.imtc-files.org/jvt-experts or ftp://ftp.ietf.org/.



INTEL-1117
9,641,841

outside the scope of the JVT standardization. The NAL decoder interface is normatively defined in the JVT coding standard, whereas the interface between the VCL and the NAL is conceptual and helps in describing and separating the tasks of the VCL and the NAL. The VCL specifies an efficient representation for the coded video signal. The NAL abstracts the VCL from the details of the transport layer used to carry the VCL data. It defines a generic and network independent representation for information above the level of the slice. Both, VCL and NAL are media-aware, i.e. they may know properties and constraints of the underlying networks, such as the prevailing or expected packet loss rate, maximum trnasfer unit (MTU) size, and transmission delay jitter.

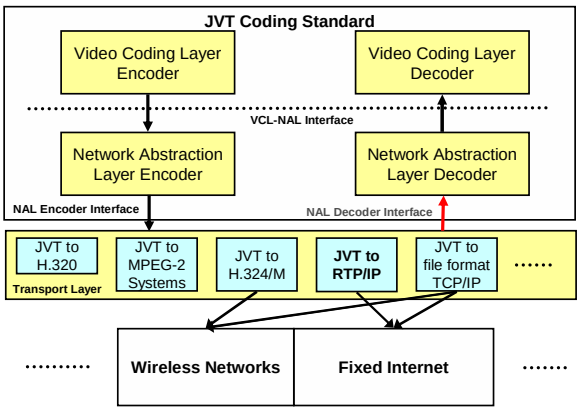


Figure 1: The JVT Standard in Transport Environment

Transport protocols are very heterogeneous in terms of reliability, Quality-of-Service guarantees, encapsulation, and timing support. Moreover, transport systems differ in terms of internal setup and configuration protocol availability. For conversational applications, the usage of setup and configuration protocols for capability exchange, such as SIP/SDP [22], [23] for IP-based applications and H.245 in H.324 systems [24], is relatively common. For broadcast and multicast applications the session announcement protocol (SAP) [22] is usually applied or systems specifications like MPEG-2 systems [6] define appropriate announcement messages. The mapping specification of JVT coded data to transport protocols is completely outside of the JVT standardization effort. However, the NAL concept provides on one hand significant flexibility to integrate JVT coded data into existing and future networks, and, on the other hand, it also takes care of maintaining a sufficient common basis to facilitate gateway design between different transport protocols.

### 2.2 Video Coding Layer – Compression Tools

Although the design of the JVT codec basically follows the design of prior video coding standards such as MPEG-2, H.263, and MPEG-4, it contains many new features that enable it to achieve a significant improvement in terms of compression efficiency. We will briefly highlight those, for more details we refer to [1], [2] and [4]. In the following we describe working draft version 2 (WD-2) [1] in greater detail as the software used in the experiments reflects the status of WD-2. We will briefly present major changes from WD-2 to CD-1 at the end of this section. In JVT coding according to WD-2, blocks of 4x4 samples are used for transform coding, and thus a macroblock (MB) consists of 16 luminance and 4 blocks for each chrominance component. Conventional picture types known as I- and P-pictures are supported. Furthermore, JVT coding supports multi-frame motion-compensated prediction (MCP). That is, more than one prior coded picture can be used as reference for the motion compensation. Encoder and decoder have to store already coded pictures in a multi-frame buffer. A generalized frame-buffering concept has been adopted allowing MCP not just from previous frames but also from future frames. For that, a flexible and efficient signaling method has been adopted. In addition, JVT coding permits generalized B-Pictures allowing two prediction signals per block but reference more than one picture. However, with appropriate multiple reference frame handling the well-known functionality of disposable B-pictures known from, e.g., MPEG-2 [5] is still supported. This allows for example temporal scalability. To simplify stream switching JVT provides S-pictures; for details and applications, see [37].

A MB can always be coded in one of several INTRA-modes. There are two classes of INTRA coding modes, one which basically allows coding flat regions with low frequency components and one which allows to code details in a very efficient way utilizing prediction in the spatial domain by referring neighboring samples of already coded blocks. In addition to the INTRA-modes, various efficient INTER-modes are specified in JVT coding. In addition to the SKIP-mode that means just copying the content from the same position from the previous picture, seven motion-compensated coding modes are available for MBs in P-pictures. Each motion-compensated mode corresponds to a specific partition of the MB into fixed size blocks used for motion description. Currently, blocks with sizes of 16x16, 16x8, 8x16, 8x8, 8x4, 4x8, and 4x4 samples are supported by the syntax, and thus up to 16 motion vectors maybe transmitted for a MB.

The JVT coding syntax supports quarter-sample accurate motion compensation. The motion vector components are differentially coded using either median or directional prediction from neighboring blocks. The chosen prediction depends on the block shape and the position inside the MB. JVT coding is basically similar to other prior coding standards in that it utilizes transform coding of the prediction error signal. However, in JVT coding the transformation is applied to 4x4 blocks and, instead of the DCT, JVT coding uses a separable integer transform similar to a 4x4 DCT. Since the inverse transform is defined by exact integer operations, inverse-transform mismatches are avoided. Appropriate transforms are used to the four DC-coefficients of each chrominance component (2x2 transform) and the INTRA16x16-mode (repeated 4x4).

For the quantization of transform coefficients, JVT coding uses scalar quantization. The quantizers are arranged in a way that there is an increase of approximately 12.5% from one quantization parameter (QP) to the next. The quantized transform coefficients are scanned in a zig-zag fashion and converted into coding symbols by run-length coding (RLC). All syntax elements of a MB including the coding symbols obtained after RLC are conveyed by entropy coding methods.

JVT coding supports two method of entropy coding. The first one called Universal Variable Length Coding (UVLC) uses one single infinite-extend codeword set. Instead of designing a different VLC table for each syntax element, only the mapping to the single UVLC table is customized according to the data statistics. The efficiency of entropy coding is improved if Context-Adaptive Binary Arithmetic Coding (CABAC) is used that allows the assignment of non-integer numbers of bits to each symbol of an alphabet. Additionally, the usage of adaptive codes permits the adjustment to non-stationary symbol statistic and context modeling allows for exploiting statistical dependencies between symbols. However, we will use the UVLC method in the remainder of this work as CABAC is likely only part of a high coding efficiency profile. For removing block-edge artifacts, the JVT coding design includes an inloop deblocking filter. The JVT coding block edge filter is applied inside the motion prediction loop. The filtering strength is adaptively controlled by the values of several syntax elements.

In CD-1 [2] the zig-zag scanning and run-length coding is replaced by context-adaptive variable length codes (CVLC). All UVLCs are replaced by CVLC which are adapted to the statistics of different syntax elements. In addition, quantizer values have been shifted. For high-complexity modes an Adaptive Block Transform (ABT) with similar properties as the 4x4 integer transform allowing different transform sizes up

to 16x16 has been introduced. However, all changes from WD-2 to CD-1 are of little relevance to the presented results and conclusions in this paper.

### 2.3 Network Abstraction Layer and IP-Based Transmission over Fixed and Wireless Networks

The Network Abstraction Layer of JVT video defines the interface between the video codec itself and the outside world. It operates on Network Abstraction Layer Units (NALUs) which give support for the packet-based approach of most existing networks. At the NAL decoder interface it is assumed that the NALUs are delivered in transmission order and that packets are either received correctly, are lost, or an error flag in the NALU header is set if the payload contains bit errors.

A NALU consists of a one-byte header and a bit string that is, in most cases, the bits representing the MBs of a slice. The header byte itself consists of the aforementioned error flag, a priority field to, e.g. signal disposable NALUs, and the NALU type. The NAL payload type either indicates the included video data type, i.e. single slice packet or one of three data partitions, or high-level information, such as random access points, parameter set information, or supplemental enhancement information. The NAL specification provides means to transport high-level syntax, i.e., syntax which is assigned to more than one slice, e.g. to a picture, a group of pictures or to an entire sequence. The applied parameter concept used in JVT is significantly different to previous video coding standards as NALUs are self-contained packets. High-level information is stored in parameter sets. Each parameter set can be transmitted in the session setup or during the session in an asynchronous and reliable way well before the synchronous video data references it. In IP environments for example, the Session Description Protocol (SDP) [23] can be used to define parameter sets that are conveyed reliably using Session Initiation Protocol (SIP) [22] or Real Time Streaming Protocol (RTSP) [26]. No redundancy coding techniques for headers are necessary, because there are no headers above the slice header due to the Parameter Set concept. The standardization process of the RTP payload specification for JVT [26] is still an ongoing process. However, the draft RTP payload specification is aligned to the goal of simple syntax specification and expects that NALUs are transmitted directly as the RTP payload except for the additional concept of aggregation packets.

For packet-switched real-time services, the third generation partnership project (3GPP) has chosen to use SIP and SDP [28] for call control and RTP for media transport [29]. Figure 2 shows the packetization of an NALU through the 3GPP user plane protocol stack. The NALU is mapped to an RTP payload according to [26]. After Robust Header Compression (RoHC) [30] this IP/UDP/RTP packet is finally encapsulated to a Radio Link Control (RLC)-SDU. If any of the RLC-PDUs containing data from a certain RLC-SDU has not been received correctly, the RLC-SDU is typically discarded. The RLC/RLP layer can and should perform re-transmissions if the application has relaxed delay constraints as typical for streaming services. However, especially in conversational applications RLC retransmissions are not feasible due to stringent delay constraints; then, an erroneous RLC frame usually results in a loss of the entire IP/RTP packet and the included NAL unit.

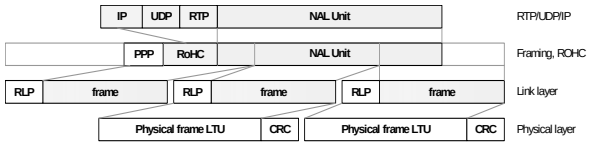


Figure 2: Packetization of NAL units through 3GPP user plane protocol stack

Usually two kinds of errors are present in today's transmission systems: bit inversion errors or packet losses. Combinations of both are also possible, especially when transmitting over heterogeneous networks including wireless networks. However, all relevant protocols or multiplexer including UDP/IP and almost all underlying mobile systems include packet loss and bit error detection capabilities applying sequence numbering and block check sequences, respectively. Therefore, it can be assumed that a vast majority of erroneous transmission packets can be detected. Some research has been conducted in decoding bit-error prone video packets, and, in very few scenarios, gains have been reported. However, in the design of the JVT video codec bit-erroneous NALUs are ignored. On the one hand, this simplifies standard design and test model software implementation. On the other hand very few transport protocols and receivers process bit-error prone packets as the expected gain is marginal compared to the associated implementation costs. However, the error indication flag in NALU header provides flexibility. In the remainder of this work we will focus on packet-lossy transmission.

### 2.4 Common IP-based Test Conditions

The JVT acknowledged the importance of IP-based transmission over fixed and wireless networks by adopting a set of common test conditions for fixed and mobile IP based transmission, in [31] and [32], respectively. These test conditions allow for selecting appropriate coding features, testing and evaluating error resilience tools, and producing meaningful anchor results. The common defined test case combinations include fixed Internet conversational services as well as packet-switched conversational services and packet-switched streaming services over 3G mobile networks. Also included is simplified offline network simulation software, which uses appropriate error patterns captured under realistic transmission conditions. Anchor video sequences, appropriate bit rates and evaluation criteria are specified. Extensive results for Internet and mobile test conditions are presented for example in [33], [34], and [35]. In the remainder we will only present results for a small but representative selection of the common Internet test conditions. The applied test case combinations include the QCIF sequences *Foreman* and *Hall Monitor* with frame rate 30 frames per second (fps). The first 300 frames of the original sequence are encoded at a frame rate of 7.5 fps for Foreman and 15 fps for Hall Monitor applying an IPPP… sequence. Although QCIF resolution will mainly be used in wireless IP-based environments, it also provides a good indication on the performance of JVT coding for fixed Internet with CIF or even higher spatial resolution. As the current JVT test model software does not include a rate control we chose to present the results when encoding the sequence with a fixed quantization parameter. For all test results we encoded the sequence with the quantization parameter $q = 12, 14, \ldots, 30$ (according to WD-2) and measured the resulting total bit rate including a 40 byte IP/UDP/RTP header for each transmitted NALU. As performance measure we chose the commonly applied averaged Peak-Signal-to-Noise ratio of the luminance component (Y-PSNR) where the average is taken over all encoded frames. For all experiments we transmitted at least 3000 frames to obtain sufficient statistics. The only applied channel error pattern is one of four different Internet error patterns captured from real-world measurements [36]. This error pattern results in a packet loss rate of approximately 10%. For simplicity we did not consider the mobile test conditions where the packet loss rate depends on the length of the packets as the probability that a short packet is hit by a bit error is lower than the loss probability for a long packet. However, the general results provided here also apply to transmission over wireless channels. Some pointers and further explanation will follow.

## 3 Error Resilience Features in JVT and IP-based Networks

### 3.1 System Overview and Problem Formulation

The investigated video transmission system is shown in Figure 3. JVT video encoding is based on a sequential encoding of frames denoted with the index $n$, $n = 1, \ldots, N$ with $N$ the total

number of frames to be encoded. In most existing video coding standards including JVT, within each frame video encoding is typically based on sequential encoding of MBs (exceptions are discussed later) denoted by index $m$, $m = 1,\ldots,M$ where $M$ specifies total number of MBs in one frame and depends on the spatial resolution of the video sequence. MBs are generally quadratic with size $\sqrt{I} \times \sqrt{I}$ pixel, i.e. one MB contains $I$ pixel and the position is denoted with $i$ where $i = 1,\ldots I$. The pixel value in the original sequence in frame $n$ and MB $m$ at MB position $i$ is denoted as $s_{n,m,i}$.

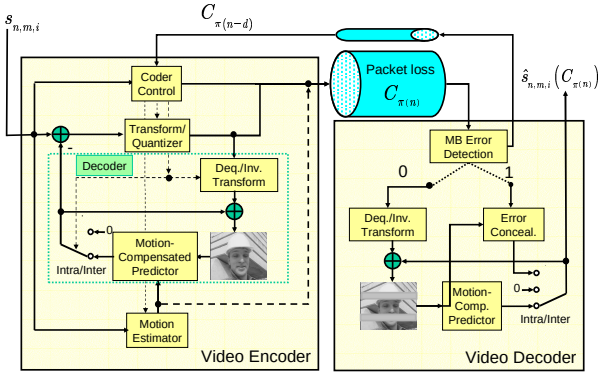


Figure 3: JVT coding in network environment with packet losses and delayed feedback information.

The generated video data is packetized and transmitted over a packet lossy channel. The encoding process can form slices by grouping a certain number of MBs. The slice within each frame is indexed by $j$ with $j = 1,\ldots,J_n$ and $J_n$ the total number of slices in frame $n$. Each slice $j$ contains a number $\tilde{m}_j$ specifying the spatial location of the first MB in this slice. The number of MBs contained in slice $j$ is defined as $\Delta m_j$. Picture number $n$ and start MB address $\tilde{m}_j$ are binary coded in the slice header. Although $n$ usually uses a predefined modulo-counter we will ignore this for ease of exposition. For our experimental results in the following we will use a packetization such that each frame is packetized into 3 slices, i.e. $\forall_{n=1,\ldots,N}\ J_n = 3$, and the number of MBs for each frame is fixed to 33, i.e. $\forall_{j=1,\ldots J_n}\ \Delta m_j = 33$.

For notational convenience, let us define the number of packets necessary to transmit all frames up to $n$ as $\pi(n) = \sum_{n'=1}^{n} J_{n'}$, with the inherent assumption that one slice is transported in one packet. With that, we can define the packet loss or channel behavior $c$ as a binary sequence $\{0,1\}^{\pi(n)}$ indicating whether a slice is lost (indicated by 1) or correctly received (indicated by 0). Obviously, if a slice is lost all MBs contained by this slice are lost. We can assume that for most transport protocols the decoder is aware of any lost packet as discussed previously. The channel loss sequence is obviously random and, therefore, we denote it as $C_{\pi(n)}$ where the statistics are in general unknown to the encoder.

The decoder processes the received sequence of packets. Whereas correctly received packets are decoded as usual for the lost packet an error concealment algorithm has to be invoked. The reconstructed pixel $\hat{s}_{n,m,i}$ at position $i$ in MB $m$ and frame $n$ depends on the channel behavior and on the decoder error concealment. In Inter mode, i.e., when MCP is utilized, the loss of information in one frame has a considerable impact on the quality of the following frames, if the concealed image content is referenced for motion compensation. Because errors remain visible for a longer period of time, the resulting artifacts are particularly annoying to end-users. Therefore, due to the motion compensation process and the resulting error propagation the reconstructed image depends not only on the lost packets for the current frame but in general on the entire channel loss sequence $C_{\pi(n)}$. We denote this dependency as $\hat{s}_{n,m,i}\left(C_{\pi(n)}\right)$.

According to Figure 3**,** in conversational applications a low bit-rate reliable back-channel from the decoder to the encoder is usually available which allows reporting a $d$-frame delayed version $C_{\pi(n-d)}$ of the observed channel behavior at the decoder to the encoder. In IP environments for example, this can be based on RTCP messages. Details on this feedback exploitation will be discussed in Section 5.

From this system perspective an error-resilient video coding standard suitable for conversational IP-based services has to provide features to combat various problems, always focusing on prime goal of high compression efficiency. The tools required in error-prone environment can be divided into two major categories according to the problem to be solved: On the one hand it is necessary to avoid errors completely and to minimize the visual effect of errors within one frame. On the other hand, as errors cannot be avoided, the well-known problem of spatio-temporal error propagation in hybrid video coding has to be limited. In the following we will discuss JVT standard features and test model extensions for encoder and decoder which address solution to the discussed problems.

### 3.2 Error Resilience Features in JVT

Packet loss probability and the visual degradation from packet losses can be reduced by introducing slice-structured coding, which provides spatially-distinct resynchronization points within the video data for a single frame. Slices also provide syntactical resynchronization points. All predictions related to the entropy coding are reset at the start of a slice. With that, on the one hand the packet loss probability can be reduced, if slices and, therefore, transmission packets are relatively short, since the probability of a bit error hitting a short packet is generally lower than for long packets. Moreover, short packets reduce the amount of lost information and, hence, the error is limited and error concealment methods can be applied successfully. In the JVT test model the simple previous frame copy (PFC) error concealment has been replaced by advanced error concealment (AEC) [37] which makes use of the packetized transmission. On the other hand, the loss of spatial prediction within one frame and the increased overhead associated with decreasing slices adversely affect performance. Especially for mobile transmission, where the packet length affects the loss probability of a packet a careful selection of the packet length is necessary. A detailed discussion on this issue for the JVT codec and for the IP-based 3G mobile test conditions is presented in [26]. A packet length selection in the range of around 500 bytes is suggested as satisfying compromise between overhead and resulting loss probability.

JVT coding specifies several enhanced concepts to reduce the artifacts caused by packet losses within one frame. Slices can be grouped by the use of compound packets into one packet and, therefore, concepts such as Group-of-Block (GOB) and *Slice Interleaving* [39], [40] are possible. This does not reduce the coding overhead in the VCL, but the costly IP overhead of typically 40 bytes per packet can be avoided. A more advanced and generalized concept is provided by *Flexible MB Ordering* (FMO) [35], [41] which has been introduced in order to have the possibility to transmit MBs in non-scan order. This flexibility allows the definition of different patterns including slice interleaving without interrupting the inter MB prediction for motion vector prediction and entropy coding. FMO is especially powerful with appropriate error concealment. A third error resilience concept included in JVT is *data partitioning*, which can also reduce visual artifacts resulting from packet losses, especially if prioritization or unequal error protection is provided by the network. For more details on the data-partitioning mode we refer to [1] and [4]. In general any kind of *forward error protection* (FEC) in combination with interleaving for packet lossy channels can be applied. A simple solution is provided by RFC2733 [42], more advanced schemes have been evaluated in many papers, e.g. [43], [44]. However, in the following we do not consider FEC schemes in the transport layer as this requires a reasonable number of packets per codeword.

For conversational applications, however, the number of packets per channel codeword should be low to avoid overhead from transport protocols and reduced coding efficiency as well as to limit the delay. In addition, short codes are generally not very powerful which limits the applicability to the investigated low-delay applications.

Despite all these techniques, packet losses and resulting reference frame mismatches between encoder and decoder are usually not avoidable. Then, the effects of spatio-temporal error propagation are in general severe. The impairment caused by transmission errors decays over time to some extent. However, the leakage in standardized video decoders, such as JVT, is not very strong, and quick recovery can only be achieved when image regions are encoded in Intra mode, i.e., without reference to a previously coded frame. Completely Intra coded frames are usually not inserted in real-time and conversational video applications as the instantaneous bit rate and the resulting delay is increased significantly. Instead, JVT coding allows encoding of single MBs for regions that cannot be predicted efficiently as it is also known from other standards. Another feature in JVT is the possibility to select the reference frame from the multi-frame buffer. Both features have mainly been introduced for improved coding efficiency, but they can efficiently be used to limit the error propagation. Conservative approaches transmit a number of Intra coded MBs anticipating transmission errors. In this situation, the selection of Intra coded MBs can be done either randomly or preferably in a certain update pattern. For details and early work on this subject see [45], [46] and [47]. Multiple reference frames can also be used to limit the error propagation, for example in *video redundancy coding* schemes (see e.g. [48]). In addition, a method known from H.263 under the acronym *redundant slices* will be supported in JVT coding. This will allow sending the same slice predicted from different reference frames which provides the decoder the possibility to predict this slice from error-free reference areas. Finally, multiple reference frames can be successfully combined with a feedback channel. This will be discussed in detail in Section 5.

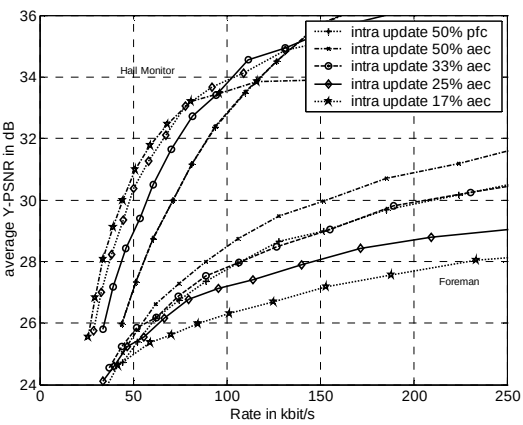


Figure 4: Rate-Distortion performance of simple error resilience in JVT coding: different error concealment and pseudo-random intra updates with varying update frequency for Hall Monitor and Foreman.

Figure 4 shows the performance of simple error resilience tools in JVT coding according to specified conditions as defined in section 2.4 and 3.2 for Hall Monitor and Foreman. Different error concealment strategies are assessed and the results show that already for moderate movement (Foreman) a significant gain for AEC compared to the PFC is visible. In addition, different pseudo-random intra-update ratios are evaluated. From the results it is obvious that an appropriate intra-update ratio can increase the overall quality significantly. However, it is also apparent that optimal update ratio not only depends on the channel statistics, but also on the transmitted sequence and possibly even on the transmission bit rate. For Foreman, 50% intra update shows the best performance, whereas for Hall Monitor in general less intra update is recommended, especially for lower bit rates. Simple adaptive coding schemes have been presented in [45], [46] and [47]. However, these heuristic approaches can not fully exploit the optimal performance. In the following we present methods which aim to optimize rate-distortion performance for packet-lossy channels.

## 4 Encoder Enhancements – Adaptive MB Intra Updates

### 4.1 Rate-Distortion Optimized Mode Selection

JVT coding consists of a motion-compensation and a residual coding stage. The task of residual coding is to refine signal parts that are not sufficiently well represented by motion-compensated prediction. From the viewpoint of bit allocation strategies, the various modes relate to various bit rate partitions. The concept of selecting appropriate coding options in many source-coding standards is based on rate-distortion optimization algorithms [49], [50]. The two cost terms "rate" and "distortion" are linearly combined and the mode is selected such that the total cost is minimized. This can be formalized by defining the set of selectable coding options for one MB as $\mathfrak{O}$. In hybrid video coding systems the MB mode can be selected from the set of MB modes $\mathfrak{M}$. In the following, we assume that we only transmit one I-picture at the beginning of the video sequence and P-pictures for the remainder. However, the presented algorithm can be extended easily to other picture types. Therefore, we assume that the set of MB modes consists of two subsets, one including MB modes, which employ temporal prediction, denoted as $\mathfrak{M}_{\mathrm{P}}$ and one including pure intra coding without any prediction denoted as $\mathfrak{M}_{\mathrm{I}}$. Obviously, for I-pictures the MB mode can only be selected from $\mathfrak{M}_{\mathrm{I}}$. In JVT, not only the mode of the MB can be selected but also the reference frame can be chosen from the set of accessible reference frames $\mathfrak{R}$ [51]. The cardinality of set of reference frames $|\mathfrak{R}|$ specifies the maximum number of reference frames. The set of accessible coding options for P-frames is defined as all possible combinations of MB modes and reference frames, i.e. $\mathfrak{O} = \left\{\mathfrak{M}_{\mathrm{I}}, \mathfrak{M}_{\mathrm{P}} \times \mathfrak{R}\right\}$. Therefore, rate-constrained mode decision selects the coding option $o^*_{n,m}$ for MB $m$ in frame $n$ such that the Lagrangian cost functional is minimized, i.e.

$$o^*_{n,m} = \arg\min_{o \in \mathfrak{O}} \left( D_{n,m}(o) + \lambda R_{n,m}(o) \right). \tag{1}$$

In the JVT test model for coding efficiency, the distortion $D_{n,m}(o)$ is the sum of squared pixel differences (SSD), i.e.

$$D_{n,m}(o) = \sum_{i=1}^{I} \left| s_{n,m,i} - \hat{s}_{n,m,i}(o) \right|^2, \tag{2}$$

where $\hat{s}_{n,m,i}(o)$ is the reconstructed pixel value at the decoder in frame $n$ and MB $m$ at position $i$ when encoding with mode $o$. The rate $R_{n,m}(o)$ is simply obtained by encoding with mode $o$ and the Lagrange parameter is selected as $\lambda = C_\lambda \cdot 2^{q/3}$ with $C_\lambda = 0.85$ [52].

In case of error prone transmission we like to replace the distortion in (2) with a more meaningful measure. Assuming that the encoder is aware of the channel statistics $C_{\pi(n)}$, the encoder can get an estimate of the reconstructed pixel values at the decoder by the expected distortion as

$$D_{n,m}\left(o, C_{\pi(n)}\right) = \sum_{i=1}^{I} E_{C_{\pi(n)}} \left| s_{n,m,i} - \hat{s}_{n,m,i}\left(o, C_{\pi(n)}\right) \right|^2, \tag{3}$$

where the expectation is over the channel $C_{\pi(n)}$. In the following we will discuss and assess different possibilities to obtain an estimation of the decoder distortion. We will also address the assumption on the availability of the channel statistics at the encoder.

### 4.2 Estimation of Decoder Distortion

The estimate of the expected pixel distortion in packet loss environment has been addressed in several previous papers. For example, in [53] or [54] models to estimate the distortion introduced due the transmission errors and the resulting drift are defined. A similar approach has recently been proposed within the JVT project which attempts to measure the drift noise between encoder and decoder [55]. In all these approaches the quantization noise and the distortion introduced by the transmission errors are linearly combined. The encoder keeps track of an estimated pixel distortion and, therefore, requires additional complexity in encoder. The addition is approximately one-time the decoder complexity as for each pixel the drift noise has to be computed and stored. An accurate estimation of the expected pixel value at the decoder for H.263-like coding can be achieved by the recursive optimal per-pixel estimate (ROPE) algorithm [56]. ROPE provides an accurate estimation by keeping track of the first and second order moment of $\hat{s}_{n,m,i}$, $E\left\{\hat{s}_{n,m,i}\right\}$ and $E\left\{\hat{s}^2_{n,m,i}\right\}$, respectively. As two moments for each pixel have to be tracked in the encoder the added complexity of ROPE is approximately twice the complexity of the decoder. However, the extension of the ROPE algorithm to JVT coding is not straight-forward. The in-loop filter, the sub-pel motion accuracy and the advanced error concealment require taking into account the expectation of products of pixels at different positions to obtain an accurate estimation which makes the ROPE either infeasible or inaccurate in this case.

Therefore, a powerful yet complex method has been introduced into the JVT test model to estimate the expected decoder distortion [57]. The encoder obtains an estimate of the reconstructed value at the decoder and, therefore, of the expected distortion as

$$D_{n,m}\left(o, C_{\pi(n)}\right) = \sum_{i=1}^{I} E_{C_{\pi(n)}} \left| s_{n,m,i} - \hat{s}_{n,m,i}\left(o, C_{\pi(n)}\right)\right|^2, \qquad (3)$$

where the expectation is over the channel $C_{\pi(n)}$. Let us assume that we have $K$ copies of the random variable channel behavior at the encoder, denoted as $C_{\pi(n)}(k)$. Additionally, assume that the set of random variables $C_{\pi(n)}(k)$, $k = 1, \ldots, K$ are *identically* and *independently* distributed (*iid*). Then, as $K \to \infty$, it follows by the strong law of large numbers that

$$\frac{1}{K}\sum_{k=1}^{K}\left|s_{n,m,i} - \hat{s}_{n,m,i}\left(C_{\pi(n)}(k)\right)\right|^2 = E_{C_{\pi(n)}}\left|s_{n,m,i} - \hat{s}_{n,m,i}\left(C_{\pi(n)}\right)\right|^2, \qquad (4)$$

holds with probability 1. An interpretation of the left hand side leads to a simple solution of the previously stated problem to estimate the expected pixel distortion. In the encoder $K$ copies of the random variable channel behavior and the decoder are operated. The reconstruction of the pixel value depends on the channel behavior $C_{\pi(n)}(k)$ and the decoder including error concealment. The $K$ copies of channel and decoder pairs in the encoder operate independently. Therefore, the expected distortion at the decoder can be estimated accurately in the encoder if $K$ is chosen large enough. However, the added complexity in the encoder is obviously at least $K$ times the decoder complexity.

### 4.3 Implementation Aspects and Performance Comparison

In [57] it was shown that the mode selection for packet lossy channels with packet loss probability $p$ can be carried out according to

$$o^*_{n,m} = \arg\min_{o \in \mathfrak{O}} \left(\hat{D}_{n,m}(o) + \hat{C}_\lambda 2^{q/3} R_{n,m}(o)\right), \qquad (5)$$

where $\hat{D}_{n,m}(o)$ defines the expected distortion for MB $m$ in frame $n$ assuming that all transmission packets of frame $n$ are received correctly, but the reference frames are erroneous based on the random packet loss sequence $C_{\pi(n-1)}$. Additionally, it was shown in that the Lagrange parameter should be adapted to a value $\hat{C}_\lambda \le C_\lambda$ depending on the selected mode and the loss rate. However, the benefits of adaptation are marginal and, therefore, due to simplicity it is proposed to set $\hat{C}_\lambda = C_\lambda$. For the error-robust MB mode and reference frame selection in frame $n$ we therefore encode each MB $m$ with each accessible MB mode $o \in \mathfrak{O}$. Then, for each combination $(n, m, o)$ the expected distortion is estimated by using either the drift noise estimation, the ROPE algorithm or multiple decoders (MD) with $K = 100$. For the ROPE only the closest ful-pel motion vector position is used in the update process for the first and second order moment and the loop-filter operation is ignored. In the following we assume for all methods a statistically independent packet loss probability of 10% at the encoder and the simple PFC error concealment in the distortion estimation. However, at the decoder the AEC is applied and the channel according to the error pattern described in section 2.4 is used.

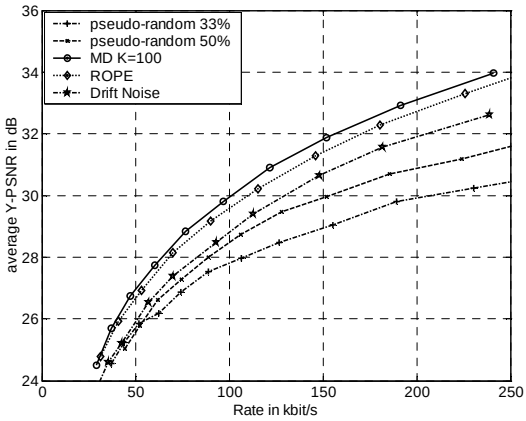


Figure 5: Rate-Distortion performance (Foreman) for different adaptive MB mode and reference frame selection compared to pseudo-random updates.

Figure 5 shows rate-distortion performance (Foreman) for different adaptive MB mode and reference frame selection schemes compared to pseudo-random updates. It can be observed that all channel and content-adaptive mode selection modes outperform the best regular intra-update strategies. However, it can also be observed that the quality of the estimated decoder distortion significantly influences the performance. Comparing the gains of the multiple decoders with the best regular intra updates a gain of about 1-2.5 dB depending on the bit rate is obvious. For the same quality the bit-rate decreases for adaptive intra-updates by about 30%. The ROPE compared to the optimized multiple decoder distortion in general has higher bit rate and higher average PSNR for the same quantization parameter. This means that ROPE estimates the decoder distortion in general too high resulting from the fact of ignoring loop filter and fractional-pel motion compensation. The drift noise estimation method generates similar bit rates as the multiple decoder approach, however, the placement of the intra refresh is not optimal. From a complexity-performance point of view the multiple decoder approach is not suitable, but ROPE provides excellent results with acceptable encoder complexity even in real-time encoding. A better adjustment of ROPE to JVT coding as well as to the advanced error concealment in the JVT test model is subject of future work. However, for full exploitation of the performance of JVT we use the multiple decoder approach in the encoder for the remainder of this work.

For all the previous investigations we assumed that the channel statistics or at least the average packet loss rate is known at the encoder. Obviously, this is in general not the case or the estimation is not accurate. To evaluate the stability, but also the optimality of the channel-adaptive approaches, encoding at different expected error rates has been performed, while the transmission scenario is not altered. Without showing the detailed results it is worth to mention that loss prob-

ability estimation errors in the range of halving or doubling the error rate results in negligible overall performance loss. In addition the optimality when knowing the exact error rate at the encoder could be verified.

## 5 Exploiting Network Feedback in Video Encoders

### 5.1 Overview

So far we have assumed that there is no feedback information from the decoder except for a possible report of an average packet loss rate. However, as already mentioned in Section 3, the knowledge of a $d$ -frame delayed version of the observed channel characteristic $C_{\pi(n-d)}$ at the encoder might be useful even if the erroneous frame has already been decoded and presented. This characteristic can be conveyed from the decoder to the encoder by acknowledging correctly received slices (ACK), sending a not-acknowledge message (NACK) for missing slices or both types of messages. In general it can be assumed that the reverse channel is error-free and the overhead is negligible. The feedback can be used to limit the error propagation. In the following we will discuss several scenarios and show the performance of selected results. Most of the techniques rely on an appropriate selection of the reference frames or the insertion of intra information. As JVT coding allows selecting intra updates and reference frames on MB basis a combination with the optimized MB mode selection and reference frame selection according to Section 4.1 is appropriate.

A simple yet powerful approach suitable for video codecs using just one reference frame such as MPEG-2, H.261 or H.263 version 1 has been introduced in [58] and [59] under the acronym *Error Tracking*. When receiving a NACK on parts of frame $n-d$ or the entire frame $n-d$ , the encoder attempts to track the error to obtain an estimate of the quality of frame $n-1$ , which serves as reference for frame $n$ . Having tracked the error, the encoder can perform one of the following three options: a) the MBs in frame $n$ that would have referenced a damaged area are coded in intra mode; b) the referencing is only restricted to non-damaged areas; c) the same type of error concealment is performed in the encoder as the decoder would apply for frames $n-d, n-d+1, \ldots, n-1$ such that the reference frames in encoder and decoder match. Whereas a) and b) are rather straightforward in implementation and can be used in combination, c) is more difficult as encoder and decoder have to apply the identical error concealment, see e.g. [58], [59], [60], and [61]. Note that with this concept error propagation in frame $n$ is only removed if frames $n-d+1, \ldots, n-1$ have been received at the decoder without any error. We will discuss these issues in further detail when adapting the presented methods to JVT coding.

A technique addressing the problem of continuing error propagation has been introduced, among others, in [62], [63], and [64] under the acronym *NEWPRED*. Based on these early non-standard compliant solutions in H.263 Annex N [8] a reference picture selection (RPS) for each GOB is specified such that the NEWPRED technique can be applied. RPS can be operated in two different modes. In the negative acknowledgement mode (NAM) the encoder only alters its operation in the case of reception of a NACK. Then, the encoder attempts to use an intact reference frame for the erroneous GOBs. To completely eliminate error propagation this mode has to be combined with independent segment decoding (ISD) according to Annex R of H.263 [8]. In the positive acknowledgement mode (PAM) the encoder is only allowed to reference confirmed GOBs as reference. If no GOBs are available to be referenced, intra coding has to be applied. PAM and NAM could be combined in a similar way as explained in Mode c) for the Error Tracking case by using the identical error concealment on negative acknowledged GOBs as the decoder would use. This would completely eliminate error propagation in frame $n$ even if additional errors have occurred in frames $n-d+1, \ldots, n-1$ .

### 5.2 Feedback in JVT coding – Concepts and Experimental Results

The flexibility provided in H.263 Annex U [8] and JVT coding to select the MB mode and reference frames on MB basis allows incorporating NEWPRED PAM and NAM in a straight-forward manner [54]. We will discuss two modes in the following, one based on PAM only and the second based on PAM and NAM. In the case of PAM only the encoder is only allowed to reference acknowledged area. The MB mode and reference frame selection is performed according to the description in Section 4.1, with the modification that certain areas from some reference frames are restricted. In general this allows complete removal of mismatch - independent of the applied decoder error concealment. However, as JVT coding applies a deblocking filter operation in the motion compensation loop over slice boundaries, a complete removal of encoder and decoder mismatch is not possible but the influence of this mismatch is negligible.

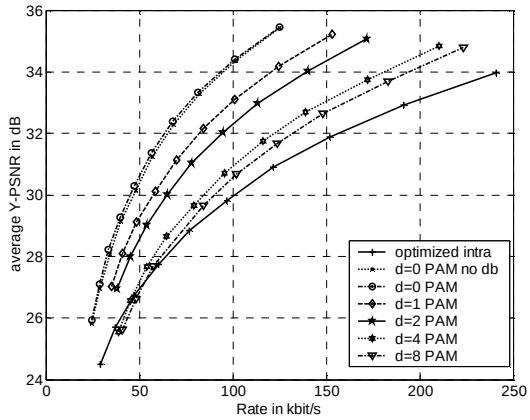


Figure 6: Rate-Distortion performance (Foreman) for positive acknowledge mode (PAM) only for different frame delays and deblocking filter modes compared to optimized intra updates applying AEC.

Figure 6 shows rate-distortion performance (Foreman) for PAM only for different frame delays and deblocking filter modes compared to optimized intra updates applying AEC at the receiver. Note that a frame delay of $d$ results in feedback delay of $d/7.5\,\text{sec}$ . The results show that for any delay this system outperforms the best system without any feedback using optimized intra-updates. For small delays, the gains are significant and for the same average PSNR the bit-rate is less than 50%. With increasing delay the gains are reduced, but compared with the highly complex mode decision without feedback this method is still very attractive. Obviously, this high-delay result is strongly sequence dependent but for other sequences similar results have been verified. An additional advantage of the PAM results from the fact that the encoder does have to be aware of the applied error concealment in the encoder as long as correctly received pixels are not altered significantly. The figure shows also that the influence of loop filter mismatch is less significant than the loss of coding efficiency when turning off the loop filter for the entire sequence. Adaptive switching of the loop filter is subject of further study within the JVT project.

In a second mode PAM and NAM are combined such that the encoder reconstructs the identical reference frames as the decoder using the identical error concealment. Only completely reconstructed frames are referenced in the investigated system. The selection of the appropriate MB mode and reference frames is again based on the concept in Section 4.1, but the reference frames are now possibly error-prone. The rate-distortion optimization takes care of selecting the appropriate MB mode, i.e. intra mode insertion, or the appropriate reference frame for each MB.

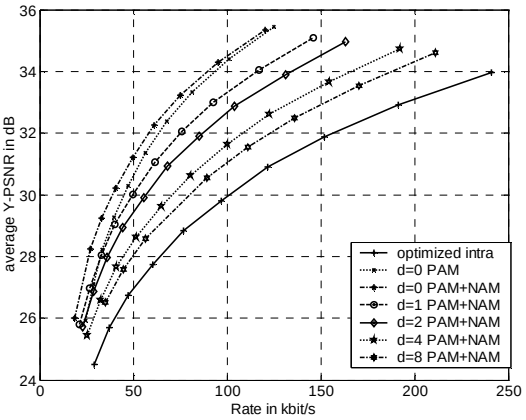


Figure 7: Rate-Distortion performance (Foreman) for combined PAM and NAM for different frame delays compared to PAM only and optimized intra updates; AEC is applied for all cases at encoder at decoder.

Figure 7 shows rate-distortion performance (Foreman) for combined PAM and NAM for different frame delays compared to PAM only and optimized intra updates; AEC is applied for all feedback cases at encoder and decoder. The results for this mode applying both PAM and NAM show similar results as PAM only. The performance for increasing delay decreases in a similar way as it does for the PAM only. Whereas for low bit rates combined PAM and NAM shows some gain compared to PAM only, for higher bit rates the performance difference almost vanishes. This results from the fact that for low bit rates referencing a concealed area is often significantly better in terms of rate-distortion performance than coding this area in intra mode. However, for high bit rates bad reference frames are not used and the intra mode is selected more often by the rate-distortion optimized mode selection. The little gains combined with the disadvantage of necessary normative error concealment for the combined PAM and NAM makes the PAM only preferable in practicable systems. The loop filter problem causing small mismatch between encoder and decoder is currently under discussion in the JVT project as it also of importance for other applications. The final syntax specification will probably allow turning on and off this filter for each slice.

Recently, combination of adaptive mode selection and feedback-based drift removal methods have been proposed in [51] and [54]. Although these methods involve some additional complexity as the statistical estimation of reference frames has to be adapted for each received feedback information, it was shown that especially for medium to higher feedback delays these methods can provide significant gains. The application to JVT coding is subject of future work.

## 6 Conclusions

In this work we have applied widely accepted standard-compliant techniques to enhance the quality of JVT coded video transmitted over packet lossy networks. A summary of the results is shown in Figure 8. The macroblock mode and the reference frame selection are extended to include the expected decoder distortion in the Lagrangian mode decision. This increases the quality of video in packet-lossy IP environment significantly as shown in the diagram. In addition, the exploitation of network feedback has been studied and several schemes and their dependency on the feedback delay have been assessed. All presented feedback-based schemes enhance the quality of the decoded video significantly, even for moderate and higher delays of about 1 second (see d=8). From a system point-of-view the best performance is provided by the PAM only mode which does not require standardized error concealment and still provides almost identical performance as the combination of PAM and NAM. PAM only is superior to simple PFC error concealment in combination with combined PAM and NAM.

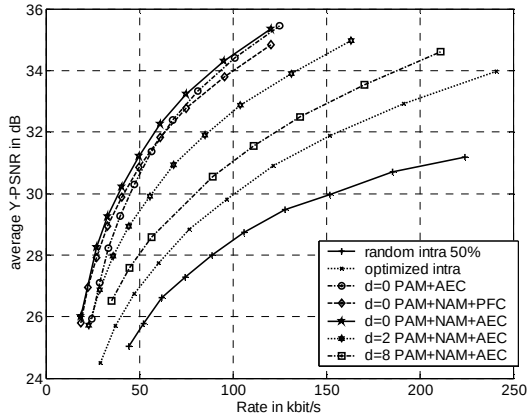


Figure 8: Comparison of rate-distortion performance (Foreman) of different investigated transmission schemes.

Future work for improved error resilience of JVT coding in conversational IP-based applications includes the combination of macroblock mode selection and network feedback exploitation as well as the combination of the presented methods with FMO, data partitioning and FEC. As streaming applications delay constraints are more relaxed, the importance of link layer or transport layer retransmission protocols become more important. These issues in combination with appropriate encoding methods, e.g. multi-frame handling or S-picture functionalities, and appropriate buffer management (see e.g. [65]) are subject of ongoing and future research.

## Acknowledgements

The authors would like to thank the diploma students Tongmin Xu and Florian Obermeier who implemented and tested parts of the presented algorithms. They would also thank Thomas Wiegand, Miska Hannuksela, and Gary Sullivan for ongoing and valuable discussions on these subjects within and outside of JVT. Finally, the authors would like to thank Prof. Girod for the invitation to this excellent workshop.

## References

[1] T. Wiegand (ed.), “Working Draft Number 2, Revision 4 (WD-2),” JVT-B118r7, Apr. 2002.

[2] T. Wiegand (ed.), “Committee Draft Number 1, Revision 0 (CD-1),” JVT-C167, May 2002.

[3] A. Joch, F. Kossentini, P. Nasiopoulos,” A Performance Analysis of the ITU-T Draft H.26L Video Coding Standard”, Proc. Packet Video Workshop, Apr. 2002.

[4] G. Sullivan and T. Wiegand (ed.), “Special Issue on H.26L/JVT Coding”, IEEE CSVT, in preparation, Oct. 2002.

[5] J. Postel, “Internet Protocol”, RFC 791, Sep. 1981.

[6] ISO/IEC International Standard 13818; “Generic coding of moving pictures and associated audio information”, Nov. 1994.

[7] ISO/IEC JTC1, “Generic Coding of Audiovisual Objects – Part 2: Visual (MPEG-4 Visual)”, ISO/IEC 14496-2, Version 1: Jan. 1999, Version 2: Jan. 2000; Version 3: Jan. 2001.

[8] ITU-T Recommendation H.263, “Video Coding for Low Bit-Rate Communication”, Version 1: Nov. 1995, Version 2: Jan. 1998, Version 3: Nov. 2000.

[9] T. Socolofsky and C. Kale, “A TCP/IP Tutorial,” RFC 1180, Jan. 1991.

[10] J. Postel, “User Datagram Protocol”, RFC 768, Aug. 1980.

[11] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson, “RTP: A Transport Protocol for Real-Time Applications”, RFC 1889, Jan. 1996.

[12] D. Hoffman, G. Fernando, V. Goyal, R. Civanlar, “RTP Payload Format for MPEG1/MPEG2 Video,” RFC 2250, Jan. 1998.

[13] Y. Kikuchi, T. Nomura, S. Fukunaga, Y. Matsui, H. Kimata, “RTP Payload Format for MPEG-4 Audio/Visual Streams”, RFC3016, Nov. 2002.

[14] C. Bormann, L. Cline, G. Deisher, T. Gardos, C. Maciocco, D. Newell, J. Ott, G. Sullivan, S. Wenger, C. Zhu, “RTP Payload Format for the 1998 Version of ITU-T Rec. H.263 Video (H.263+)”, RFC2429, Oct. 1998.

[15] P. Bahl and B. Girod (Eds.), “Wireless Video”, vol. 36, IEEE Communcations Magazine, June 1998.

[16] J. C. Brailean, T. Sikora, and T. Miki (Eds.), “Special Issue on Error Resilience”, vol. 14, Signal Processing: Image Communication, May 1999.

[17] H. Gharavi and L. Hanzo (Eds.), “Special Issue on Video Transmission for Mobile Applications”, Proc. IEEE, Oct. 1999.

[18] A. Reibman and M. T. Sun (Eds.), “Wireless Video”, Marcel Dekker, 2000.

[19] C. W. Chen, P. Cosman, N. Kingsbury, J. Liang, and J. W. Modestino (Eds.), “Special Issue on Error-Resilient Image and Video Transmission”, IEEE JSAC, vol. 18, no. 6, June 2000.

[20] Y. Wang, S. Wenger, J. Wen, and A. K. Katsaggelos, „Error Resilient Video Coding Techniques“, IEEE Signal Processing Magazine, pp. 61-82, July 2000.

[21] ITU-T Recommendation H.320, “Narrow-Band Visual Telephone Systems and Terminal Equipment,” Rev. 4 May 1999.

[22] M. Handley, H.Schulzrinne, E.Schooler, and J.Rosenberg, "SIP: Session Initiation Protocol", RFC 2543, March 1999.

[23] M. Handley, V. Jacobson, “SDP: Session Description Protocol”, RFC 2327, April 1998.

[24] D. Lindbergh, “The H.324 Multimedia Communication Standard,“ IEEE Comm. Mag., vol. 34, no. 12, pp. 46-51, Dec. 1996.

[25] M. Handley, C. Perkins, and E. Whelan, “Session Announcement Protocol,” RFC 2974, Oct. 2000.

[26] A. Rao and R. Lanphier, “Real Time Streaming Protocol (RTSP)”, RFC 2326, Apr. 1998.

[27] S. Wenger, T. Stockhammer, M. Hannuksela, “RTP payload Format for JVT Video”, draft-wenger-avt-rtp-jvt-01.txt, Internet Draft, Work in Progress, July 2002.

[28] 3GPP Technical Specification 3GPP TS 24.229: “3rd GPP; Technical Specification Group Core Network; IP Multimedia Call Control Protocol based on SIP and SDP.”

[29] 3GPP Techn. Specification 3GPP TS 26.235: “Packet switched conversational multimedia applications; Default codecs.”

[30] Hannu, H., Jonsson, L-E., Hakenberg, R., Koren, T., Le, K., Liu, Z., Martensson, A., Miyazaki, A., Svanbro, K., Wiebke, T., Yoshimura, T. and H. Zheng, “RObust Header Compression (ROHC): Framework and four profiles: RTP, UDP, ESP, and uncompressed,” RFC 3095, July 2001.

[31] S. Wenger, “Common Conditions for wireline, low delay IP/UDP/RTP packet loss resilient testing”, ITU-T SG16 Doc. VCEG-N79r1, Sept. 2001.

[32] G. Roth, R. Sjöberg, G. Liebl, T. Stockhammer, V. Varsa, and M. Karczewicz, “Common Test Conditions for RTP/IP over 3GPP/3GPP2,” ITU-T SG16 Doc. VCEG-N80, Sept. 2001.

[33] T. Stockhammer, M. Hannuksela, and T. Wiegand, “JVT/H.26L in 3G Wireless Environments”, submitted for IEEE CSVT, Special Issue on H.26L/JVT Coding, April 2002.

[34] T. Stockhammer, T. Wiegand and S. Wenger, “Optimized transmission of H.26L/JVT coded video over packet-lossy networks,” Proc. ICIP 2002, Rochester, NY, Sept. 2002.

[35] S. Wenger and M. Horowitz, “Scattered Slices: Simulation Results,” JVT-C090, May 2002.

[36] S. Wenger, “Error Patterns for Internet Experiments,” ITU-T SG16 Doc. Q15-I-16r1, October 1999.

[37] R. Kurceren and M. Karczewicz, “A Proposal for SP-frames,” ITU-T SG 16 Doc. VCEG-L27, Eibsee, Germany, Jan. 2001.

[38] V. Varsa, M.M. Hannuksela, and Y. Wang, “Non-normative error concealment algorithms,” ITU-T VCEG-N62, Sept. 2001.

[39] S. Wenger and G. Cotê, “Using RFC 2429 and H.263+ at low to medium bit-rates for low-latency applications”, Proc. Packet Video Workshop 1999, New York, NY, April 1999.

[40] V. Varsa and M. Karczewicz, “Slice interleaving in compressed video packetization”, Proc. Packet Video Workshop 2000, Forte Village, Italy, May 2000.

[41] S. Wenger and M. Horowitz, “Flexible MB Ordering – A New Error Resilience Tool for IP-Based Video”, Proc. IWDC 2002, Capri, Italy, Sept. 2002.

[42] J. Rosenberg and H. Schulzrine, “An RTP Payload Format for Generic Forward Error Correction,” RFC 2733, December 1999.

[43] G. Carle, and E.W. Biersack, “Survey of Error Recovery Techniques for IP-based Audio-Visual Multicast Applications,” IEEE Network Magazine, Vol. 11, No. 6, p. 2-14, Nov. 1997.

[44] U. Horn, K. Stuhlmüller, M. Link, and B. Girod, “Robust Internet Video Transmission based on Scalable Coding and Unequal Error Protection”, IEEE Trans. on Image Processing, vol. 15, no. 1-2, pp. 77-94, Sept. 1999.

[45] Q. F. Zhu and L. Kerofsky, “Joint Source Coding, Transport Processing, and Error Concealment for H.323-Based Packet Video,” Proc. SPIE VCIP, vol. 3653, pp. 52–62, Jan. 1999.

[46] P. Haskell and D. Messerschmitt, “Resynchronization of Motion-Compensated Video Affected by ATM Cell Loss,” Proc. IEEE ICASSP, vol. 3, pp. 545–548, 1992.

[47] J. Liao and J. Villasenor, “Adaptive Intra Update for Video Coding over Noisy Channels,” Proc. ICIP, vol. 3, pp. 763–766, Oct. 1996.

[48] S. Wenger, G. Knorr, J. Ott, and F. Kossentini, “Error resilience support in H.263+,” IEEE CSVT, pp. 867-877, November 1998.

[49] T. Wiegand, M. Lightstone, D. Mukherjee, T. G. Campbell, and S. K. Mitra. “Rate-Distortion Optimized Mode Selection for Very Low Bit Rate Video Coding and the Emerging H.263 Standard,” IEEE CSVT, 6(2):182–190, April 1996.

[50] G.J. Sullivan and T. Wiegand, “Rate-Distortion Optimization for Video Compression,” IEEE Signal Processing Magazine, vol. 15, no. 6, pp. 74-90, Nov. 1998.

[51] T. Wiegand and B. Girod, “Multi-frame Motion-Compensated Prediction for Video Transmission,” Kluwer Academic Publishers, Sept. 2001.

[52] T. Wiegand, and B. Girod, “Lagrangian Multiplier Selection in Hybrid Video Coder Control,” Proc. ICIP 2001, Oct. 2001.

[53] G. Cote, S. Shirani, F. Kossentini, “Optimal mode selection and synchronization for robust video communications over error-prone networks,” IEEE JSAC, vol. 18, no. 6, pp. 952 -965.

[54] T. Wiegand, N. Färber, K. Stuhlmüller, and B. Girod, “Error-resilient video transmission using long-term memory motion-compensated prediction”, IEEE JSAC, vol. 18, no. 6, pp. 1050-1062.

[55] C. W. Kim, D. W. Kang, and I. S. Kwang, “High-Complexity Mode Decision for Error Prone Environment”, JVT-C101, May 2002.

[56] R. Zhang, S. L. Regunathan, and K. Rose, “Video Coding with Optimal Inter/Intra-Mode Switching for Packet Loss Resilience,” in IEEE JSAC, vol. 18, no. 6, pp. 966-976.

[57] T Stockhammer, D. Kontopodis, and T. Wiegand, “Rate-Distortion Optimization for JVT/H.26L Coding in Packet Loss Environment”, Proc. Packet Video Workshop 2002, April 2002.

[58] E. Steinbach, N. Färber, and B. Girod, “Standard Compatible Extension of H.263 for Robust Video Transmission in Mobile Environments,” IEEE CSVT, vol. 7, no. 6, pp. 872-881, Dec. 1997.

[59] B. Girod and N. Färber, “Feedback-Based Error Control for Mobile Video Transmission,” Proceedings of IEEE, vol. 97, no. 10, pp. 1707-1723, Oct. 1999.

[60] W. Wada, “Selective Recovery of video packet losses using error concealment”, IEEE JSAC, vol. 7 , pp. 807-814, June 1989.

[61] Y. Wang, and Q. Zhu, “Error control and concealment for video communication: a review,” Proceedings of IEEE, vol.86, pp-974-997, May 1998.

[62] T. Nakai, and Y. Tomita, “Core experiments on feedback channel operation for H.263+,” ITU-T SG15 LBC96-308, Nov. 1996.

[63] S. Fukunaga, T. Nakai, and H. Inoue, ” Error resilient video coding by dynamic replacing of reference pictures,” Proc. IEEE Globecom, vol.3, Nov. 1996.

[64] Y. Tomita, T. Kimura, and T. Ichikawa, “Error resilient modified inter-frame coding system for limited reference picture memories,” Proc. PCS, Sept. 1997.

[65] J. Cabrera, A. Ortega, and J. I. Ronda, “Stochastic Rate-Control of Video Coders for Wireless Channels,” IEEE CSVT, vol. 12, no. 6, pp. 496-510, June 2002.

Corpus ID: 16634279

Share

# Standard-Compliant Enhancements of JVT Coded Video for Transmission over Fixed and Wireless

T. Stockhammer, S. Wenger • Published 2002 • Computer Science, Engineering

This paper describes standard compliant enhancements of the currently developed JVT coding algorithms for transmission over fixed and wireless IP-based networks. This includes a description of JVT coding algorithm and the adaptation to IP-based networks. The error resilience features within the JVT coding algorithm are presented in greater detail. Standard compliant encoder and decoder enhancements as well as the exploitation of network information to improve the quality in packet lossy… Expand

9 Citations
Highly Influential Citations 2
Background Citations 4
Methods Citations 4
View All

No Paper Link Available

Save
Alert

Figures
Topics
9 Citations
104 References
Related Papers

## Figures from this paper

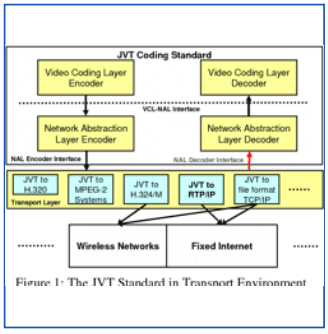


Figure 1

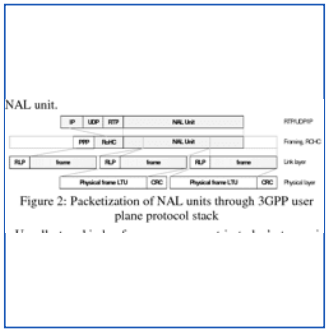


Figure 2

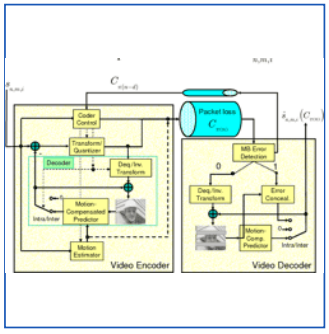


Figure 3

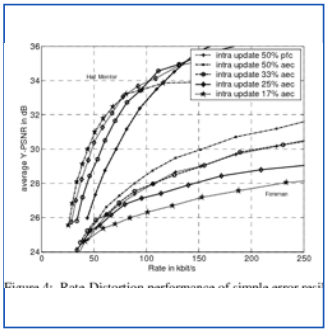


Figure 4

View All 8 Figures & Tables

## Topics

AI-Generated

Joint Video Team | Error-Resilience | Video Coding Standards

## 9 Citations

Sear | Citation Type | Has PDF | Author | More Filters | More Filters | Filters | Sort by Relevance

### Color-Aware Packet Marking Based on H.264AVC Coder Extensions for Improved Video Quality

S. Przylucki  Computer Science, Engineering · 2014

TLDR Attention is focused on the extensions of the H.264 standard that increase the error-resilience, particularly Data Partitioning (DP) and Flexible Macroblock Ordering (FMO). Expand

Highly Influenced | PDF  8 Excerpts  Save

### AVC in Wireless Environments

T. Stockhammer  M. Hannuksela  T. Wiegand  Computer Science, Engineering

TLDR The main objective of this paper is to provide an overview over the tools which are likely to be used in wireless environments and discusses the most challenging application, wireless conversational services in greater detail. Expand

32 | PDF  Save

### Cross-layer feedback-based resilient coding for video transmission over uplink cellular network

Chongyang Zhang  Hua Yang  Songyu Yu  Xiangzhong Fang  Computer Science, Engineering ·

IEEE transactions on consumer electronics · 2007

TLDR A novel error resilient video coding scheme for the uplink video applications over cellular networks is proposed that adopts a radio link layer agent at the sender to implement the cross-layer feedback. Expand

1 | Highly Influenced | PDF  8 Excerpts  Save

### A highly efficient, low delay architecture for transporting H.264 video over wireless channel

Hong-Bin Yu  Songyu Yu  Ci Wang  Computer Science · Signal processing. Image communication · 2004

9  3 Excerpts  Save

### Wine2 wireless network demonstration platform for ip-based real-time multimedia transmission

T. Stockhammer  G. Liebl  H. Jenkac  P. Strasser  Daniel Pfeifer  J. Hagenauer  Computer Science, Engineering ·

2003

TLDR A flexible and fast, but yet accurate simulation environment for wireless systems, which satisfies real-time constraints imposed by the respective applications and protocols of interest, is presented and the suitability of this environment to test, evaluate, and assess video-conferencing applications is shown. Expand

7

Save

### Application of binning codes to video coding

J. Chou | A. Tabatabai | Mohammed Zubair Visharam | Computer Science ·

Visual Communications and Image Processing · 2005

TLDR This paper proposes a method that incorporates binning codes and channel decoding to address video coding problems that have multiple possible predictions at the decoder and shows that the results are superior to pre-existing methods. Expand

3 | Save

### Isolated regions in video coding

M. Hannuksela | Ye-Kui Wang | M. Gabbouj | Computer Science · IEEE transactions on multimedia · 2004

TLDR This paper proposes the use of the isolated regions coding tool that jointly limits in-picture prediction and interprediction on a region-of-interest basis and can be applied as an error-robust macroblock mode decision method and used in combination with unequal error protection. Expand

79 | PDF | 1 Excerpt | Save

### Interactive Error Control for Mobile Video Telephony

Waqar Zia | Tauseef Afzal | Wen Xu | G. Liebl | T. Stockhammer | Computer Science, Engineering ·

IEEE International Conference on Communications · 2007

TLDR This work shows how a low-complexity interactive error tracking technique can be combined with multiple reference picture selection (RPS) based on the existing syntax of H.264/AVC to outperforms other interactive error protection strategies for moderate channel loss rates, with minimal impact on end-to-end delays. Expand

7 | Highly Influenced | PDF | 3 Excerpts | Save

### Subjective Evaluation of Error Control Strategies for Mobile Video Communication

Waqar Zia | Tobias Oelbaum | K. Diepold | Computer Science, Engineering · Picture Coding Symposium · 2007

TLDR Effective error control techniques suitable for H.264/AVC in such a complexity constrained conversational environment are proposed and evaluated by extensive subjective testing to ensure an enhanced quality of service (QoS). Expand

2 | PDF | 1 Excerpt | Save

## 104 References

Sear | Citation Type | Has PDF | Author | More Filters | More Filters | Filters | Sort by Relevance

### Optimized transmission of H.26L/JVT coded video over packet-lossy networks

T. Stockhammer | T. Wiegand | S. Wenger | Computer Science, Engineering ·

Proceedings. International Conference on Image... · 2002

TLDR A flexible and robust rate-distortion optimization technique is presented to select coding mode and reference frame for each macroblock and derives a method to obtain an estimate of the decoder pixel distortion at the encoder. Expand

39 PDF Save

---

### Error resilient video coding techniques

Yao Wang S. Wenger Jiantao Wen A. Katsaggelos Computer Science, Engineering · IEEE Signal Processing Magazine · 2000

TLDR The majority of the article is devoted to the techniques developed for block-based hybrid coders using motion-compensated prediction and transform coding, and a separate section covers error resilience techniques for shape coding in MPEG-4. Expand

569 Save

---

### Robust Internet video transmission based on scalable coding and unequal error protection

U. Horn Klaus Stuhlmüller M. Link B. Girod Computer Science, Engineering · Signal processing. Image communication · 1999

298 Save

---

### Optimal mode selection and synchronization for robust video communications over error-prone networks

G. Côté S. Shirani F. Kossentini Computer Science, Engineering · IEEE Journal on Selected Areas in Communications · 2000

We describe an effective method for increasing error resilience of video transmission over bit error prone networks. Rate-distortion optimized mode selection and synchronization marker insertion… Expand

233 PDF Save

---

### A Performance Analysis of the ITU-T Draft H . 26 L Video Coding Standard

Anthony Joch F. Kossentini P. Nasiopoulos Computer Science, Engineering · 2002

TLDR A detailed analysis of the individual coding gains provided by several key features of the H.26L standard can be used to motivate further improvements to the standard, assist in choosing the features that will comprise each of its profiles, and facilitate the development efficient encoding algorithms that find an optimal balance between coding performance and complexity. Expand

36 PDF Save

---

### Error resilient video coding by dynamic replacing of reference pictures

S. Fukunaga T. Nakai H. Inoue Computer Science · Proceedings of GLOBECOM'96. 1996 IEEE Global… · 1996

TLDR The proposed error resilient video coding system, which dynamically replaces reference pictures in inter-frame coding according to the acknowledgement signals from the decoder, outperforms NACK signaling especially under the existence of the backward channel errors which are highly correlated with the forward channel errors. Expand

84 Save

---

### Rate-Distortion Optimization for JVT/H.26L Video Coding in Packet Loss Environment

T. Stockhammer Dimitrios Kontopodis T. Wiegand Computer Science, Engineering · 2002

TLDR A flexible and robust rate-distortion optimization technique for JVT/H.26L Coding is introduced and a new method on how to choose the Lagrange parameter in packet loss environments when using a rate-control is discussed. Expand

85 PDF Save

### Adaptive intra update for video coding over noisy channels

J. Liao J. Villasenor Computer Science, Engineering · Proceedings of IEEE international conference on... · 1996

TLDR An algorithm for increasing error robustness in an H.263 video coding system in the absence of a feedback channel is proposed and demonstrates a significant improvement over non-adaptive intra update strategies. Expand

74 PDF Save

### Feedback-based error control for mobile video transmission

B. Girod Nikolaus Färber Computer Science, Engineering · Proceedings of the IEEE · 1999

TLDR Feedback-based low bit-rate video coding techniques for robust transmission in mobile multimedia networks, applicable to a wide variety of interframe video schemes, including various video coding standards are reviewed. Expand

301 PDF Save

### Standard compatible extension of H.263 for robust video transmission in mobile environments

E. Steinbach Nikolaus Färber B. Girod Computer Science, Engineering · IEEE Trans. Circuits Syst. Video Technol. · 1997

TLDR The problem of robust video transmission in error prone environments is addressed utilizing a feedback channel between transmitter and receiver carrying acknowledgment information and a low complexity algorithm for real-time reconstruction of spatio-temporal error propagation is described in detail. Expand

184 Save

← 1 2 3 4 5 ... →

## Related Papers

Stay Connected With Semantic Scholar

Your E-mail Address

Sign Up

What Is Semantic Scholar?

Semantic Scholar is a free, AI-powered research tool for scientific literature, based at the Allen Institute for AI.

Learn More

About

About Us
Meet the Team
Publishers
Blog
AI2 Careers

Product

Product Overview
Semantic Reader
Scholar's Hub
Beta Program
Release Notes

API

API Overview
API Tutorials
API Documentation
API Gallery

Research

Publications
Researchers
Research Careers
Prototypes
Resources

Help

FAQ
Librarians
Tutorials
Contact

Proudly built by AI2

Collaborators & Attributions • Terms of Service • Privacy Policy • API License Agreement

AI2 Allen Institute for AI