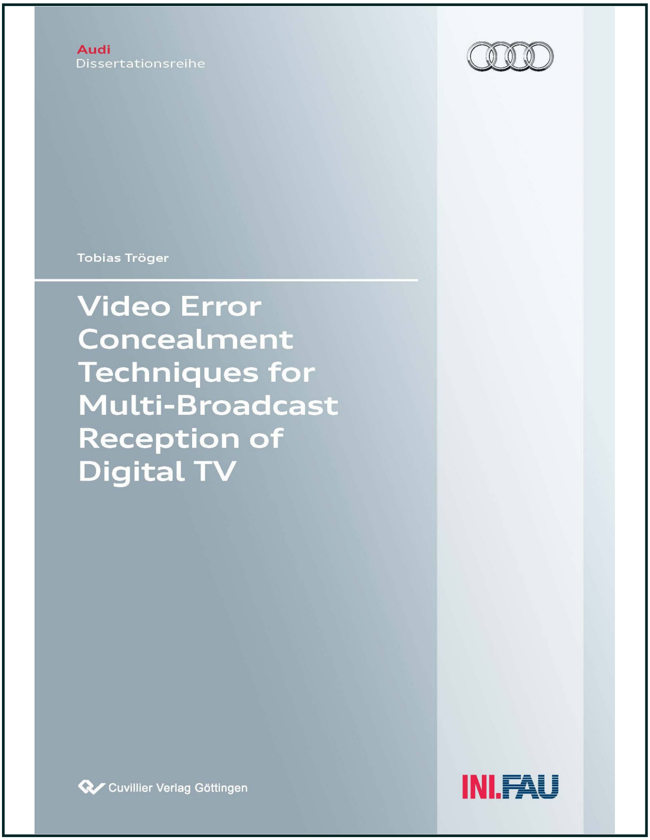
Tobias Tröger (Autor)

Video Error Concealment Techniques for Multi-Broadcast Reception of Digital TV



https://cuvillier.de/de/shop/publications/6187

1 Introduction

Since its introduction in the early 1990s, digital TV has become an essential medium deeply affecting people's everyday lives. Due to the great success, more than one billion digital TV receivers have been already sold [Scr10]. After digitization, *terrestrial* broadcasting particularly experienced a global renaissance as free-to-air TV is often enabled and mobile reception is possible in general. Providing only moderate error robustness, however, most terrestrial broadcasting standards such as ATSC, DTMB, DVB-T, and ISDB-T focus on stationary reception with home TV sets and high-mounted rooftop antennas. To ensure high visual qualities and to account for the large screen sizes of stationary TV sets, the video signals are only slightly compressed and high image resolutions are used.

Recently, people have been eager to watch digital TV not only at home but also outside. Meanwhile, terrestrial TV receivers have been integrated into cell phones, cars, and public transport vehicles. Due to fast movement and low antenna positions, however, the received signals may be shifted in frequency and shaded by hills, trees, or houses. Therefore, several terrestrial broadcasting standards such as ATSC-M/H, CMMB, DVB-H, 1Seg, and T-DMB have been tailored to mobile reception. Although enhanced error protection schemes are provided, two drawbacks can be observed. First, nation-wide network coverage is often not guaranteed. Second, the visual qualities are often low as heavy video compression is applied to enable large numbers of broadcasted TV services. Also, low image resolutions are used as the reception with small-screen devices is assumed.

As a result of the two drawbacks, many mobile devices focus on stationary TV standards although robustness is not ensured. Examples are in-car TV receivers and handheld devices such as cell phones (e.g., LG HB620T), navigation systems (e.g., Mio Moov Spirit V735), portable TV sets (e.g., Hyundai HM-T4300E), and external TV receivers for PCs (e.g., Fujitsu Slim Mobile USB DVB-T). Due to insufficient error protection, mobile TV reception with these devices typically leads to lost image slices (i.e., horizontal lines of consecutive image blocks which can not be decoded). Because of motion compensation, the losses may even propagate from the current frame into predictively-coded frames.

Aiming at a maximum viewing experience, lost slices of erroneously decoded TV signals have to be filled by *Error Concealment* (EC) techniques. In general, EC can be understood as the task to approximate the original image information as accurately as possible while also accounting for seamless integration of the filled blocks into error-free areas. To this end, classical EC techniques exploit either spatial, temporal, or spatio-temporal signal correlations. However, the reconstruction qualities are often not convincing yet.

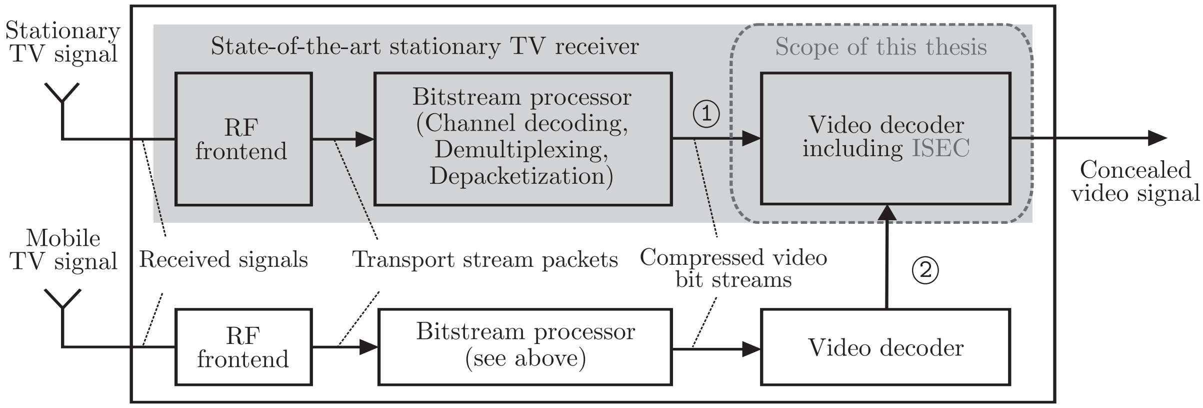


Figure 1.1: Multi-broadcast receiver utilizing both stationary and mobile TV standards. During video decoding, ISEC replaces lost slices of the high-resolution video signal ① by slices of the upscaled low-resolution reference signal ②.

To tackle low viewing experiences in mobile scenarios, *Multi-Broadcast Receivers* (MBRs) which utilize both a stationary TV standard and a mobile TV standard are currently developed. At these devices, an erroneous high-resolution video signal (representing stationary TV) and a correctly received low-resolution video signal (representing mobile TV) are available in general. As both signals represent the same TV services, upscaled images of the low-resolution video signal can be displayed if the high-resolution video signal is heavily impaired by slice losses. In the future, MBRs will be therefore integral parts of automotive infotainment systems in China (DTMB and CMMB), Europe (DVB-T and DVB-H), and the USA (ATSC and ATSC-M/H), to name only a few. Also, integration into handheld devices and portable TV sets is very likely.

Summing up, classical EC of stationary TV signals leads to low reconstruction qualities. At MBRs, mobile TV signals can be displayed alternatively. However, the visual quality is often low on large screens. This thesis therefore introduces the novel category of *Inter-Sequence Error Concealment* (ISEC) algorithms to be implemented into MBRs which utilize both stationary and mobile TV standards (see Fig. 1.1). The basic idea is to replace lost slices of the erroneous high-resolution video signal by slices of the correctly received and upscaled low-resolution reference signal during video decoding. Due to cropped and delayed images of the reference signals, robust spatio-temporal image alignment is the crucial point of ISEC. To obtain high reconstruction qualities even in case of massive compression of the reference signals, ISEC has to be extended by classical EC techniques. In addition to the outlined scenario, ISEC can also be applied to distorted scalable video signals [SMW07]. As the low-resolution reference signal in the base layer is always a purely downscaled and synchronized version of the high-resolution video signal in the enhancement layer, however, spatio-temporal image alignment can be omitted.

A special scenario for multi-broadcast reception is enabled by state-of-the-art in-car diversity receivers solely aiming at one stationary TV standard. In multi-frequency networks, such devices may receive two different physical representations of a stationary TV signal. Consequently, two erroneous high-resolution video signals are available for each TV service. Adapting the concept of ISEC to this scenario, lost blocks of the first high-resolution video signal may be replaced from the second one. Blocks which are lost in both video signals have to be concealed by classical EC techniques. While temporal alignment of both signals is still necessary, spatial alignment can be omitted. The adapted ISEC scheme may also be applied to home TV sets including digital TV tuners for cable, satellite, and terrestrial reception (e.g., triple tuners for the European DVB-C, DVB-S, and DVB-T standard which all use high image resolutions).

This thesis is organized as follows: In Chapter 2, the fundamentals of digital terrestrial broadcasting are reviewed by separately introducing two groups of TV standards which are designed for either stationary or mobile reception. For both groups, the development histories and the distribution areas are addressed. In addition, the general architectures of the corresponding broadcasting systems are studied. This includes hybrid video coding, the packetization of coded video bit streams, multiplexing with other packetized audio, video, and data streams as well as error protection by channel coding. The chapter concludes with a comparative summary of stationary and mobile TV standards.

Terrestrial broadcasting of digital TV signals over error-prone channels is described in Chapter 3. First, typical channel degradations and their effects on packetized transport streams are addressed. Second, the detection of transmission errors is described. Third, the consequences for compressed video signals are studied. Fourth, error resilience mechanisms in video communications are presented. Finally, all relevant state-of-the-art EC techniques are briefly reviewed. Three categories, namely spatial, temporal, and spatio-temporal approaches, are distinguished and the particular drawbacks are highlighted.

As classical EC methods often suffer from low reconstruction qualities, the novel ISEC scheme utilizing low-resolution reference signals is introduced in Chapter 4. Initially, automotive multi-broadcast reception of stationary and mobile TV signals is described as the main application scenario. Next, the basic concept of ISEC is introduced and the potential gain is evaluated with respect to classical EC. As the images of the low-resolution reference signal are often cropped, spatial alignment with the high-resolution video signal is addressed subsequently. After a mathematical problem formulation, pixel-based alignment exploiting a numerical optimization technique and feature-based alignment utilizing scale-invariant features are introduced. Finally, both approaches are enlarged by a temporal alignment scheme to cope with delayed reference signals. The proposed ISEC schemes incorporating spatio-temporal image alignment are extensively discussed in terms of reconstruction quality, visual quality, and computational complexity. This includes a performance comparison with classical EC techniques, an evaluation of the reconstruction quality in case of large loss areas, and a study of the influence of hybrid video coding.

In Chapter 5, algorithmic enhancements of ISEC are introduced. Aiming at an increased robustness of feature-based alignment, competing feature transforms and least-squares model fitting techniques are tested. To tackle the problem of spatio-temporal misalignment in general, two temporal adaptation strategies are also proposed. The idea is to reuse more accurate alignment parameters of previous frames in the pixel-based approach as well as the feature-based approach. Finally, enhanced interpolation schemes are studied for upscaling of the reference signal. The primary goal is to increase the maximum reconstruction quality of ISEC in case of perfect image alignment.

To guarantee high reconstruction qualities also for very low image qualities of the reference signals, *Joint Temporal and Inter-Sequence Error Concealment* (JT-ISEC) extends ISEC by a classical temporal EC technique in Chapter 6. After a short motivation, two JT-ISEC methods are proposed, especially focusing on the basic concept, the utilized temporal EC technique, and two mode selection schemes. Both JT-ISEC methods are discussed in terms of reconstruction quality and visual quality. Several extensions, either aiming at enhanced motion estimation or the restoration of the displaced frame differences, are studied and the particular quality gains are incrementally evaluated.

As the aligned low-resolution reference signals generally lack high spatial frequencies, ISEC and JT-ISEC often suffer from blurring. To recover missing spectral components, *Spatial Refinement* (SR) by frequency selective post processing is introduced in Chapter 7. First, SR is applied to ISEC. After a short motivation, the basic concept is presented and two SR techniques are proposed. The first one jointly approximates concealed blocks and neighboring correct blocks. The second one extrapolates difference image patches into the areas of the concealed blocks. Both approaches are discussed in terms of reconstruction quality, visual quality, and computational complexity. In the second part, SR is included into the concept of JT-ISEC. In doing so, the most enhanced EC scheme of this thesis is presented. Finally, its quality gain is estimated with respect to classical EC.

Chapter 8 adapts the initially proposed ISEC scheme to erroneous high-resolution reference signals. First, automotive multi-broadcast reception of stationary TV signals is described as the main application scenario. It is enabled by state-of-the-art diversity receivers in multi-frequency networks. Second, the adapted concept of ISEC is presented. While spatial alignment is omitted, classical EC of blocks which are lost in both video signals is a novel aspect. Third, drift compensation for compressed video is introduced. Finally, the reconstruction quality, the visual quality, and the computational complexity are evaluated. Here, further extensions such as JT-ISEC or SR are not necessary as the reference signal is characterized by high image resolution and high image quality.

This thesis is concluded with a summary of the main results in Chapter 9. Also, an outlook to further research topics in the field of ISEC and its extensions is included.

2 Digital Terrestrial Television – Fundamentals and Standards

In the early twentieth century, terrestrial broadcasting of analog TV evolved, enabling the reception of multimedia services at home. While analog TV has been accepted for several decades, the steadily increasing desire for better image qualities and higher image resolutions led to the transition to digital TV during the 1990s. Within the last twenty years, several competing terrestrial broadcasting techniques for digital TV have been set as standards due to technical, organizational, and political reasons. The first digital TV standards primarily aimed at stationary reception as mobile TV receivers did not exist at that time. More recent standards mainly focus on mobile reception by applying enhanced error protection schemes as handheld devices have become more and more popular. While the principle architecture of digital TV broadcasting systems aiming at stationary or mobile reception is identical, the error robustness and the video characteristics of the transmitted TV signals considerably vary because of the different application scenarios.

In this chapter, the fundamentals of digital terrestrial broadcasting are outlined, distinguishing between standards for stationary and mobile TV broadcasting in Section 2.1 and Section 2.2, respectively. Both sections address the particular development histories, distribution areas, and video signal characteristics in detail. The principle system architecture is presented on example of stationary TV broadcasting. The particular system layers are described in general from top to bottom. This includes hybrid video coding, the packetization and multiplexing of encoded video bit streams, and the protection against transmission errors by channel coding. As mobile TV broadcasting systems adopt the basic architecture of stationary TV broadcasting systems, only the most important enhancements are highlighted. Finally, the main differences between stationary and mobile TV broadcasting are summarized in Section 2.3.

2.1 Stationary Television Broadcasting

Terrestrial broadcasting of digital TV signals to stationary receivers is mostly characterized by dominant line-of-sight propagation due to high-mounted rooftop antennas. Transmission errors can be therefore largely corrected by forward error correction schemes. Besides freedom from errors, the viewers expect high image qualities of the broadcasted video signals when watching digital TV with fixed large-screen receivers.

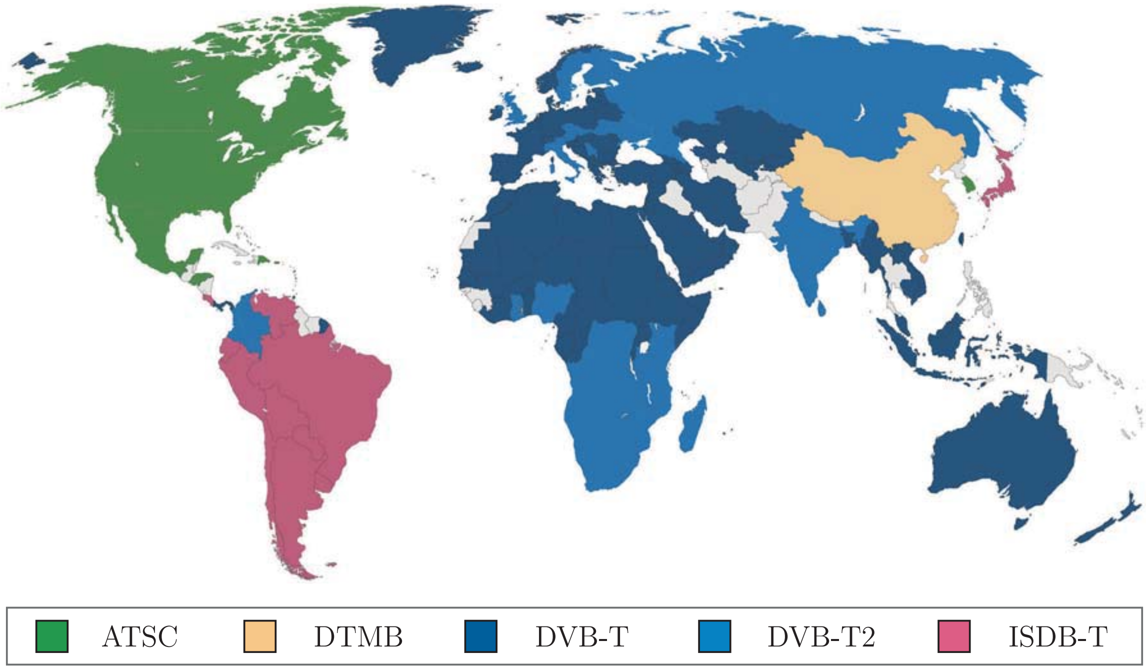


Figure 2.1: Distribution of terrestrial broadcasting standards for stationary TV reception (January 2012, [Dig12])

2.1.1 History and Distribution Areas

In 1995, the *Advanced Television Systems Committee* (ATSC) specified a standard for terrestrial transmission of digital TV signals to stationary receivers [Adv09a, RRG+06, Sgr07]. Digital TV broadcasting according to the ATSC standard was first launched in the United States in 1998. Fig. 2.1 shows that the ATSC standard is currently also established in Canada, South Korea, and some Central American countries. To compete with ATSC, the DVB group finalized the European standard *Digital Video Broadcasting – Terrestrial* (DVB-T) in 1997, meanwhile dominating digital terrestrial TV broadcasting in the world [Rei08]. The distribution area covers most European, African and Asian countries as well as Australia. Overall, DVB-T is deployed in 68 countries and adopted in 47 countries [Dig11]. With *Integrated Services Digital Broadcasting – Terrestrial* (ISDB-T), a further terrestrial broadcasting technique emerged in 1999, enabling stationary reception of digital TV [Ass05, WHRW06]. Currently, ISDB-T is deployed in Japan, and, often slightly modified, in most South American countries. For example, Brazil adopted ISDB-T under the designation *Sistema Brasileiro de Televisão Digital* (SBTVD) in late 2007. Despite the existence of the broadcasting standards ATSC, DVB-T, and ISDB-T, China set *Digital Terrestrial Multimedia Broadcast* (DTMB) as a standard by merging the two proprietary Chinese broadcasting techniques ADTB-T and DMB-T in 2006 [ZGL+07, SYY+07, Fis04]. Three years later, DVB-T2 was standardized as an enhanced version of DVB-T, introducing the latest modulation and coding techniques but still focusing on stationary TV

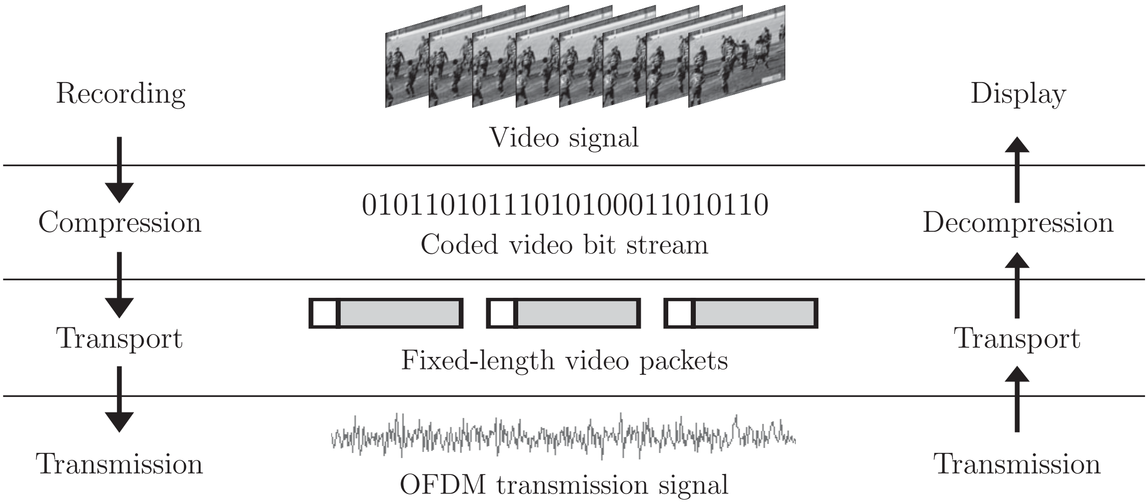


Figure 2.2: Layered architecture of digital TV broadcasting systems

broadcasting. Although DVB-T2 is still an emerging broadcasting technique, it is already deployed in Finland, Italy, Sweden, and the United Kingdom among others [Dig11].

Summing up, five terrestrial broadcasting standards exist aiming at digital TV transmission to stationary receivers. The European DVB-T standard is most widely applied, manifesting its popularity also in more than 200 million receivers being sold worldwide [Dig11]. Currently, DVB-T and the more recent DVB-T2 cover about 43% of the market of digital terrestrial TV receivers, being followed by ATSC and ISDB-T in second and third place [Scr10]. Although the number of sold DTMB receivers is the lowest at the moment, a major market growth is expected to come from DTMB over the next years.

It is worth noting that the standardization of digital terrestrial TV broadcasting is an ongoing and vital process. As novel standards like ATSC 2.0 [Win11] are currently being developed or may arise in the future, the illustrated distribution areas just represent the current situation. However, all changes are monitored in the world wide web [Tra11].

2.1.2 General System Architecture

Digital TV broadcasting has been developed for the transmission of multimedia signals being supplemented by various forms of data like commentaries, descriptive text, or files. In general, terrestrial broadcasting systems can be described by a layered architecture roughly following the well-known seven-layer *Open Systems Interconnection* (OSI) model [Zim80, Int94]. Fig. 2.2 shows the four layers which can be typically identified [RRG$^+$06]. Here, only the transmission of video signals is considered as audio signal processing is not within the scope of this thesis.

In the topmost layer, the video signals are either recorded or read from file. During recording, the video formats comprising the image resolutions and frame rates are defined.

In the subjacent compression layer, the video signals are encoded to reduce the number of bits needed for representation. Thus, the video bit rate is adjusted to the capacity of the transmission channel. The bit rate reduction is achieved by minimizing the signal redundancy as well as removing signal components which are perceptually irrelevant. During compression, the video quality of the broadcasted TV signals is determined. Typical video characteristics of stationary TV signals (i.e., video formats and coding bit rates) are discussed in Section 2.1.3. Correspondingly, hybrid video coding schemes being applied to digital TV signals are reviewed in Section 2.1.4.

In the transport layer, the coded video bit stream is segmented into packets and multiplexed with other video and audio bit streams, data files, text information such as electronic program guides, and various tables containing signaling information. Typical packetization schemes are studied in Section 2.1.5 together with packet multiplexing. After multiplexing, redundancy bits are commonly added by channel coding schemes to allow the correction of transmission errors at the receiver. Additionally, the packetized bit streams are reformatted by interleaving schemes to enhance the robustness against burst errors [Rei08]. The principles of forward error correction are reviewed in Section 2.1.6.

In the transmission layer, the bit stream is rearranged in transmission symbols which are used to modulate carrier waves. Most digital broadcasting techniques are based on multi-carrier modulation by *Orthogonal Frequency Divison Multiplexing* (OFDM) which utilizes large numbers of closely-spaced orthogonal subcarriers [AL87]. For OFDM, the bit stream is divided into several parallel substreams. Each substream is used to modulate one sub-carrier according to conventional modulation schemes such as *Quadrature Phase Shift Keying* (QPSK), 16-*Quadrature Amplitude Modulation* (16-QAM), or 64-QAM [Pro00]. In order to enable robust terrestrial broadcasting, pilot signals and guard intervals are often added prior to the digital-to-analog conversion. Transmission is mostly enabled over terrestrial channels with 6 to 8 MHz bandwidth in the *Very High Frequency* (VHF) band or the *Ultra High Frequency* (UHF) band [Int00c]. The broadcasting networks are either designed as *Single-Frequency Networks* (SFNs) where adjacent transmitters utilize the same carrier frequency, or as *Multi-Frequency Networks* (MFNs) which allow the use of different carrier frequencies [Eur08]. While broadcasting signals being transmitted in SFNs have to be identical, they may theoretically vary in MFNs due to the application of different modulation schemes.

After transmission, the four layers are passed through in bottom-up order at the receiver to recover the video and audio signals, and other data from the modulated signal (see Fig. 2.2). In detail, the receiver is tuned to the frequency of the desired broadcasting channel, the OFDM signal is demodulated, and transmission errors are corrected where possible. Subsequently, the packets belonging to the desired TV service are identified with the help of signaling information. After depacketization, the video and audio bit streams are decoded, synchronized, and sent to the display and the loudspeaker, respectively.

2.1.3 Video Characteristics

Standards for stationary TV broadcasting such as ATSC, DVB-T, or ISDB-T specify transmission systems which carry digital multimedia services and data over terrestrial channels. The generation and compression of elementary streams like digital video signals is not part of the transmission systems. In principle, most broadcasting systems are therefore open for various video formats (i.e., image resolutions and frame rates) and video coding schemes [Fis04]. However, most standardization groups have published specifications or at least implementation guidelines for the generation of standard-compliant video bit streams. In particular, the video formats and the encoding parameters are defined for widely used video coding techniques such as MPEG-2 video [Int00b] and the more recent H.264/AVC [Int05, Int10]. As the development of video coding schemes is a vital process, digital TV broadcasting systems can be expected to adopt upcoming coding techniques such as *High Efficiency Video Coding* as well [Joi10]. The following paragraphs summarize the video characteristics (i.e., the video formats and coding bit rates) that are typically observed in digital TV broadcasting to stationary receivers.

Although various video formats are allowed in stationary TV broadcasting, only high image resolutions are used in practice to account for large screens of fixed TV receivers. For example, *Standard Definition Television* (SDTV) services are either broadcasted with 720x576 pixels following the analog European PAL system, or with 720x480 pixels following the analog American NTSC system [Kum07]. The digitization of analog TV signals is specified in the widely-known ITU-R BT.601 standard [Int07b]. Alternatively, TV services may also be sent in *High Definition* (HD) with up to 1920x1080 pixels. As modern home TV sets are equipped with screens of very large sizes, these HDTV services recently have become more and more popular.

Just like the image resolutions, the frame rates of digital TV signals are often adopted from the former analog PAL and NTSC standard. Commonly, 25 frames per second (fps) or 30 fps are used for SDTV services while even up to 60 fps are typical for HDTV services.

If SDTV services are characterized by frame rates of 25 fps and image resolutions of 720x576 pixels, each being represented with a 24 bit RGB value, bit rates of approximately 250 Mbit/s result. HDTV services with 1920x1080 pixels and 60 fps even need up to 3 Gbit/s for representation with 24 bit RGB values. However, the effective data rate of broadcasting channels is typically in the order of several Mbit/s depending on transport and transmission parameters such as the channel code rate, the modulation scheme, the guard interval length, and the channel bandwidth. In Germany, the effective data rate of DVB-T channels is typically between 12 and 20 Mbit/s [Rei08]. American ATSC channels show a maximum effective data rate of 19.4 Mbit/s [RRG$^+$06].

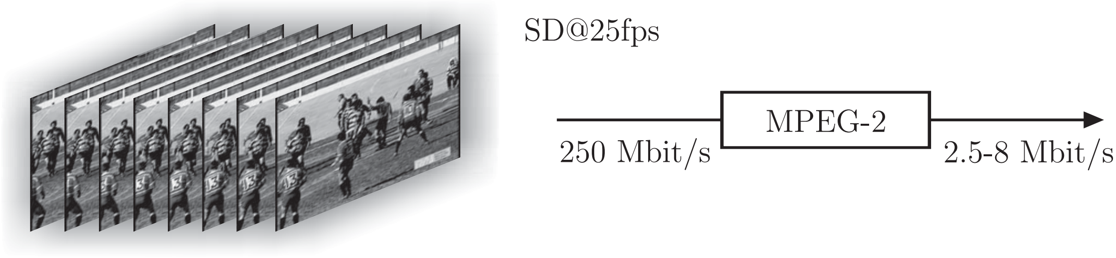


Figure 2.3: Typical video characteristics of stationary TV signals

To enable the transmission of one or more TV services over one broadcasting channel, the video bit rates have to be adjusted to the effective data rate of the broadcasting channel or to fractions of it by applying video compression. Stationary TV broadcasting according to ATSC, DVB-T, and ISDB-T is mostly based on the MPEG-2 video coding standard as it had been state-of-the-art when networks were built up and TV receivers were rolled out in the USA, Europe, and Japan. The coding bit rates of the broadcasted video signals are relatively high to enable high TV viewing experiences at home. Compressed DVB-T services typically show bit rates between 2.5 Mbit/s and 8 Mbit/s for SDTV services depending on the scene content [JDHW06, MS86]. Likewise, the bit rates of ATSC services are between 3 Mbit/s (SDTV) and 18 Mbit/s (HDTV). Consequently, only a few services are transmitted per broadcasting channel. In Germany, four DVB-T services are usually broadcasted per channel. Correspondingly, one to six services are transmitted per ATSC channel in the USA [DIF+06, Chu10]. Meanwhile, the more recent H.264/AVC video coding standard is applied to stationary TV broadcasting in several countries, enabling the transmission of HDTV with high visual quality due to enhanced coding efficiency. For example, France, Italy, and Norway use DVB-T in conjunction with H.264/AVC for the delivery of HDTV services [Dig11].

Fig. 2.3 summarizes typical video characteristics of stationary TV signals. Relatively high image resolutions like SD, moderate frame rates like 25 fps, and high image qualities due to moderate hybrid video compression according to MPEG-2 can often be observed.

2.1.4 Hybrid Video Coding

In general, compression of video signals is achieved by converting them into representations that require fewer bits than the original ones. In digital TV broadcasting, hybrid video coding is typically applied which removes spatially and temporally redundant signal information by combining transform coding and motion-compensated predictive coding [RPR77, Cla85, Gar95]. As the redundant signal parts can be fully reconstructed at the video decoder, compression is lossless up to this point. To further reduce the bit rate, however, lossy compression additionally quantizes signal information that is irrelevant to the human visual system [Ohm04].

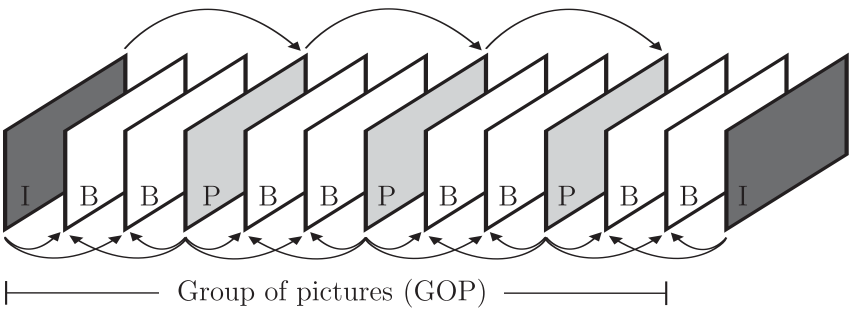


Figure 2.4: Predictive coding of B-/P-frames in hybrid video coding

Below, the basic principles of hybrid video coding are introduced mainly following the MPEG-2 video standard. The video signals are partitioned into *Groups of Pictures* (GOPs) each starting with an I-frame which is coded in *intra-mode* without temporal prediction. Only transform coding is applied. Subsequent P-frames and B-frames are coded in *inter-mode* which utilizes both transform coding and predictive coding. P-frames reference one I-frame or one P-frame by forward prediction. B-frames are bidirectionally predicted from two reference frames (either one I-frame and one P-frame, or two P-frames) applying both forward and backward prediction. Fig. 2.4 shows one GOP of an IBBP coding structure [MPFL96]. The periodical insertion of I-frames stops potential error propagation and enables random access.

The principle architecture of a hybrid video encoder is illustrated in Fig. 2.5. For block-based processing, each frame of the video signal is first segmented into non-overlapping blocks of a fixed size. In intra-mode, blocks are directly encoded by concatenating a block transform with scalar quantization and entropy coding. The prediction signal of the feedback loop is equal to zero and the prediction residual directly corresponds to the original image block. During transformation, spatial correlations between adjacent pixels are reduced and the signal energy is compacted into few transform coefficients. While MPEG-2 video coding uses the two-dimensional *Discrete Cosine Transform* (DCT) [ANR74], an integer transform being studied in [MHKK03] is utilized by the H.264/AVC standard. By quantizing the transform coefficients, irrelevant signal information is lost and thus not restorable at the decoder. Commonly, this process is controlled by a *Quantization Parameter* (QP) which scales predefined step sizes being stored in quantization tables. High QP values are equivalent to coarse quantization which leads to low coding bit rates. As the human visual system is more sensitive to low frequencies, high-frequency coefficients are typically quantized more coarsely than low-frequency ones [WZO01].

After quantization, the transform coefficients are entropy coded to remove statistical redundancies. To this end, they are re-ordered in a one-dimensional array by zig-zag scanning so that low frequency coefficients are in front of high-frequency coefficients. The latter are often zero after quantization. Run-length coding subsumes the re-ordered

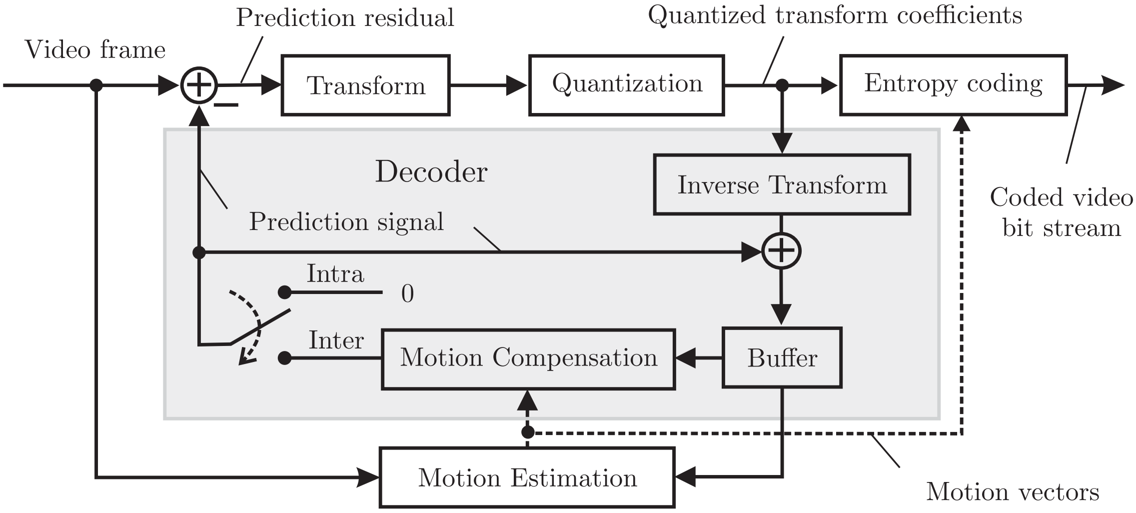


Figure 2.5: Principle architecture of a hybrid video encoder following the MPEG-2 video standard

coefficients in terms of nonzero values and the numbers of preceding zeros. Finally, entropy coding converts the different symbols, each corresponding to a pair of a zero-runlength value and a non-zero value, into binary codewords. In MPEG-2 coding, *Variable-Length Coding* (VLC) such as Huffman coding [Huf52] is preferred to fixed length coding as it is more efficient if some symbols are more likely than others [CT91]. H.264/AVC specifies context-adaptive versions of both VLC and binary arithmetic coding [MSW03].

In inter-mode, image blocks are predicted from previously coded frames which are referenced by the feedback loop of the video encoder. The loop emulates the decoder architecture to prevent drift which would occur in case of a mismatch between the encoder and the decoder (i.e., if the reference frames were not identical at the encoder and the decoder). The prediction signal is generated by motion estimation and motion compensation being typically performed on *Macroblocks* (MBs) by utilizing block matching principles [JJ81, SB00]. The *Motion Vector* (MV) represents the displacement of the MB to be encoded and the best-fitting MB of the reference frame. The prediction signal results from motion compensation (i.e., by shifting the reference block according to the determined MV). As shown in Fig. 2.5, the prediction signal which denotes the motion-compensated MB is subtracted from the MB to be encoded. The obtained prediction residual represents the information of the current MB which can not be predicted from the reference frame. It is encoded by applying a block transform, quantization, and entropy coding just like in intra mode as outlined above. For the transformation, the MBs are mostly subdivided into several smaller blocks. Entropy coding is not only applied to the quantized transform coefficients, but also to the MVs and other side information which represents picture formats and block locations.

The introduced principle of hybrid video coding is adopted by both the MPEG-2 and the H.264/AVC video standard. However, the latter introduces additional coding tools

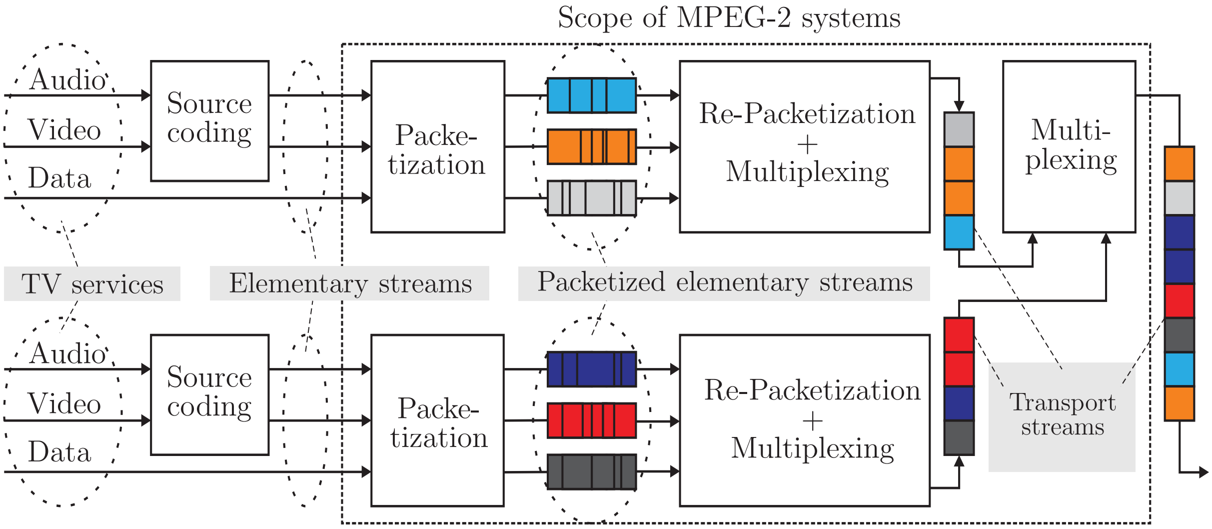


Figure 2.6: Packetization and multiplexing according to MPEG-2 systems standard

such as variable block sizes, multiple reference frames [WZG99], deblocking by in-loop filtering [LJL+03], intra frame prediction, and flexible macroblock ordering to increase the coding efficiency and the error robustness. A detailed study of the novel coding tools of H.264/AVC can be found in [WSBL03].

2.1.5 Packetization and Multiplexing

After compression based on hybrid coding schemes, the video signals are packetized for transport over unreliable terrestrial broadcasting channels. The packetized video bit stream can be multiplexed with audio bit streams and other video bit streams. For packetization and multiplexing, most digital TV standards such as ATSC, DVB-T, and ISDB-T rely on the MPEG-2 systems standard [Int07a] which specifies the generation of a *Transport Stream* (TS) consisting of small fixed-size packets. Packetization can limit the impact of transmission errors in combination with forward error correction schemes, also enabling coarse error localization in the bit stream at the receiver. Furthermore, the combination of multiple TV services in one TS is allowed due to the support of several independent time bases, namely one for every TV service.

According to the MPEG-2 systems standard, Fig. 2.6 shows the packetization and multiplexing of two digital TV services each consisting of several *Elementary Streams* (ES) such as compressed video and audio, and optional data like video text. Typically, audio and video ES are organized in fundamental coding units like encoded video frames. First, each ES is independently encapsulated into a *Packetized Elementary Stream* (PES) by accumulating an integral number of coding units. The PES packet size is variable as the data rate of compressed audio and video signals may vary over time. However, the packet size usually does not exceed 64 kbyte [Rei08]. By sharing a common timebase, synchronized decoding and presentation of video and audio PES belonging to one TV service is ensured. After packetization, all PESs of one TV service are multiplexed into a single TS. For this, the long, variable-size PES packets are segmented into short, fixed-size TS packets. In particular, each PES packet is divided into segments of 184 bytes each serving as payload of one TS packet which is supplemented by a 4-byte header. Important header elements are the *Sync Byte* defining the begin of the packet, the 1-bit *Transport Error Indicator* (TEI) flag marking packets at the receiver in case of uncorrectable transmission errors, and the 13-bit *Packet Identifier* (PID) which describes the payload [Int07a]. Finally, the two TSs each representing one TV service are multiplexed into a single TS which contains two time bases [AFZ99].

At the receiver, a particular TV service can be accessed by filtering all TS packets which contain segments of the corresponding audio, video, and data PES. The PIDs of these TS packets can be determined from the *Program Specific Information* which is regularly broadcasted within reserved TS packets in the form of tables to define the TS structure. The *Program Association Table* lists all transmitted TV services within the received TS and assigns one *Program Map Table* (PMT) to each service. Each PMT lists the PIDs of all TS packets which contain the audio, video and data PES of one particular TV service.

2.1.6 Error Protection by Channel Coding

During digital TV broadcasting over unreliable terrestrial channels, transmission errors are very likely to occur. Compressed video signals are especially prone to errors as statistical redundancies have been largely reduced. The error probability further increases due to the use of higher order modulation schemes like 64-QAM which guarantee high transmission data rates [Pro00]. Therefore, channel codes are utilized which systematically add redundancy to the TS packets in order to assist the receiver in the detection and correction of transmission errors. Most digital TV standards specify similar multi-stage channel coding schemes which combine *Block Codes* and *Convolutional Codes* [MS77]. Here, the basic principle of forward error correction is explained for the two-stage channel coding mechanisms which are used by DVB-T [Eur04, Rei08].

Fig. 2.7 shows an outer and inner channel code in conjunction with two interleaving schemes being sequentially applied to TS packets at the sender. The outer code, a (204,188)-*Reed-Solomon* (RS) code, is a systematic linear block code which is well-suited to correct burst errors (i.e., many consecutive bit errors) within TS packets [Skl88]. As bytes can be corrected independently of the number of incorrect bits, the outer code is byte-oriented. At the sender, a 16-byte checksum is calculated for each 188-byte TS packet. To this end, each byte is treated as a coefficient of a polynomial over a Galois

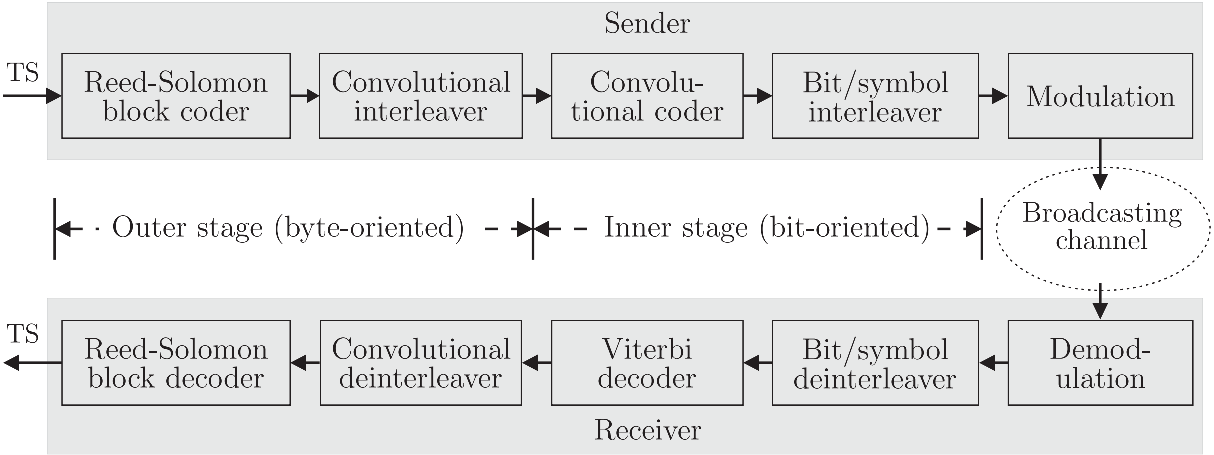


Figure 2.7: Block diagram of sender-side channel coding stage (top), and receiver-side forward error correction stage (bottom)

field [RS60]. By appending the checksum to the TS packet, it now consists of 204 bytes. At the receiver, up to 8 byte errors can be corrected within a TS packet by evaluating the checksum. If more errors occur, the RS decoder either detects a failure and marks the packet as erroneous by setting the TEI flag, or it recovers an incorrect codeword by mis-decoding without any indication. However, the probability of this decoding errors (i.e., the mapping of one RS codeword into another) is very small [MS86].

Due to the limited correction capacity, RS codes can not cope with long burst errors which result from short-term signal breakdown. Therefore, outer convolutional interleaving is applied on byte level after RS coding (see Fig. 2.7). The TS is resorted by spreading the bytes of one packet over several packets. At the receiver, outer deinterleaving breaks long burst errors within one packet into smaller ones being spread over several TS packets. Thus, the probability of successful error correction by RS decoding is enhanced [Rei08].

After outer interleaving, inner channel coding is performed. Typically, convolutional codes are applied being bit-oriented and thus well-suited to correct single bit errors. To this end, the RS-coded and interleaved TS is treated as one continuous bit stream without packet boundaries which consists of subsequent information words of k bit. For each information word, a codeword of n bit is generated by linear combination of the present k and previous m information words. Thereby, one information word is dispersed on several coding words. Practical implementations are based on shifting registers which are fed with k input sequences generating n output sequences with the help of modulo-2 adders. Often, convolutional codes are punctured to increase the code rate [KF99]. Typical (k,n,m)-triples for convolutional codes in digital TV broadcasting are (1,2,6) and (3,4,6). At the receiver, the famous Viterbi algorithm can be applied to convolutional decoding [Vit67].

After inner channel coding, inner interleaving resorts the bit stream on bit-level and also scrambles the modulation symbols, thus breaking frequency selective distortions into several single bit errors during inner deinterleaving at the receiver.

The outlined two-stage channel coding combines the effectiveness of convolutional codes against single bit errors with the high performance of RS codes in case of burst errors. At the receiver, FEC is performed as follows: Bit errors are initially corrected by convolutional decoding where possible. Afterwards, the RS decoder tries to correct the remaining incorrect bytes. Assuming that the applied convolutional code can correct one bit per RS symbol, up to 188 single bit errors can be corrected within one TS packet without utilizing the RS code. Consequently, RS decoding does not fail until 8 bytes contain more than one bit error each.

2.2 Mobile Television Broadcasting

Transmission of TV signals to mobile receivers is highly unreliable due to receiver movement and low antenna heights below ground. Even though mobile application is claimed by some broadcasting techniques which primarily aim at stationary reception (e.g., ATSC and DVB-T), error robustness is often not sufficient in practice. As a consequence, broadcasting standards which especially account for error-prone transmission to mobile devices evolved over the last years. In the following, the development history and the distribution areas of these mobile TV broadcasting standards are first summarized. Next, enhanced error protection schemes which enable an improved error robustness are described in connection with the maintained system architecture of stationary broadcasting systems. Finally, the video characteristics of mobile TV signals, such as video formats and coding bit rates, are addressed. They largely differ from those in stationary TV broadcasting as reception with small-screen devices is assumed.

2.2.1 History and Distribution Areas

During standardization of ISDB-T in 1999, a special mode being known as *One Segment* (1Seg) was defined for error-robust transmission to mobile devices [Ass05]. It uses only one of the thirteen segments of a broadcasting channel and is thus compliant to existing ISDB-T networks. Currently, mobile TV services based on the 1Seg technique are available in Japan. Some South American countries will follow due to [MS10]. Five years later, the DVB group finalized the *DVB – Handheld* (DVB-H) standard exclusively aiming at mobile TV broadcasting to mobile battery-powered devices. DVB-H is a modified version of the DVB-T standard. It provides a backwards compatible physical layer to enable the reuse of existing DVB-T networks. In 2008, DVB-H was announced as the recommended standard for mobile TV in Europe by the European Commission [Com08]. Meanwhile, DVB-H services have been commercially launched in Europe (e.g., Finland and Italy), Africa (e.g., Kenya, Namibia and South Africa) and Asia (e.g., India and Malaysia) [Dig09b, Mul10, DeR11].