# Chapter 1
# Introduction

## 1.1 Video Standard Introduction

Current video compression standards including MPEG-1/2/4 [1]–[3], H.261/2/3/4 [4]–[6], AVS [7] of China, and VC-1 [8] (an acronym for Video Codec 1 and the name of the standardized version of WMV-9) have played an important role in the world of mobile communication systems where bandwidth is still a valuable commodity. The use of modern video compression techniques offer the possibility to store or transmit the vast amount of data necessary to represent digital video in an efficient and robust way [9]. Specifically, these techniques are based on a hybrid DCT/MC coding infrastructure in Figure 1.1. They perform a block-based discrete cosine transform (DCT) to take advantage of the spatial correlation property and exploit the motion compensated (MC) prediction to improve the coding efficiency. In general, transform coding is based on dividing a frame into small blocks, taking the transform of each block, discarding high-frequency coefficients, and quantizing low-frequency coefficients. Afterward, quantized coefficients are coded using variable-length coding techniques. The size of coded streams feed back to controller so as to adjust quantization step size for achieving a target bit rate. Meanwhile, these coded streams can be stored as a digital content for video playback or sent into a wireless/broadcast channel environment for portable multimedia services. On the other hand, the decoder receives the coded video streams and performs the reverse operation to reconstruct the coded frames.

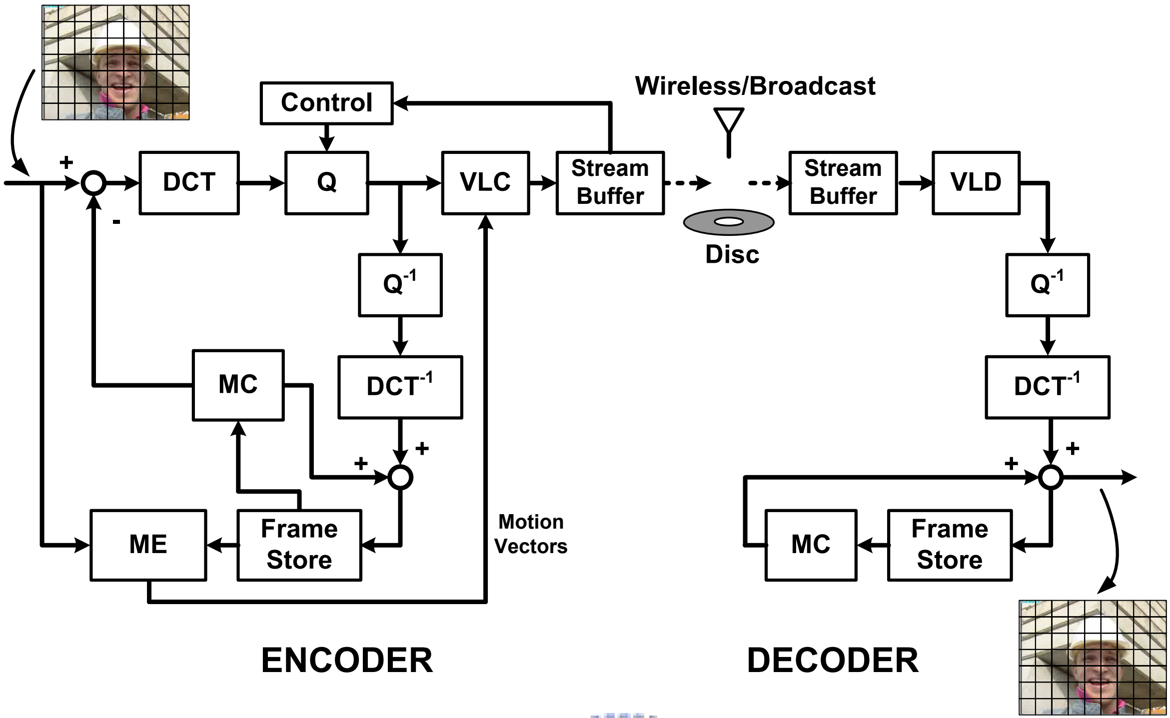



Figure 1.1: Block diagram of a basic hybrid DCT/MC coding infrastructure.

Although a wide range of video coding standards have been developed, this dissertation pays more attention on MPEG-2 [2] and H.264/AVC [6] video standards only. This is because MPEG-2 is the most widely employed today for entertainment video applications while H.264/AVC is the newest and most powerful international standard in line of the video coding standards. In general, H.264/AVC was created that improved coding efficiency by a factor of at least about two (on average) over MPEG-2 while keeping the cost within an acceptable range. However, this improved performance leads to the problem of interoperability between MPEG-2 and H.264/AVC. In other words, a video stream coded by MPEG-2 cannot be functionally decoded by H.264/AVC, and vice versa. Moreover, MPEG-2 has matured in numerous applications such as high-definition television (HDTV) and digital versatile disc (DVD), and it cannot be replaced by H.264/AVC completely even though MPEG-2 has poor performance. Hence, it is essential to jointly consider MPEG-2 and H.264/AVC for developing next-generation video engine. To this end, this dissertation first gives a high-level concept of both standards in terms of profiles and levels below and further introduces several issues for targeted mobile applications.

### 1.1.1 MPEG-2

MPEG-2 (a.k.a. H.262) is the designation for a group of coding and compression standards for Audio and Video (AV), agreed upon by MPEG (Moving Picture Experts Group), and published as the ISO/IEC 13818 international standard [2]. In general, it is based on a hybrid DCT/MC coding structure mentioned in Figure 1.1. We didn't go into the details of coding algorithm, but it can be further studied in [10][11]. Moreover, MPEG-2 is typically used to encode audio and video for broadcast signals, including direct broadcast satellite and cable TV. MPEG-2, with some modifications, is also the coding format used by standard commercial DVD movies. To clarify different applications in one coded stream, MPEG-2 limits the number of subsets of syntax by means of profiles and levels. A profile is a subset of the entire bit-stream syntax that is defined by the MPEG-2 specification [2] and listed in Table 1.1 while a level is a defined set of constraints imposed on parameters of the bit-stream in Table 1.2. For instance, the main profile and main level (MP@ML) is the most widely used for broadcast TV, and the 4:2:2 profile and main level (4:2:2@ML) is for studio video production and recording. Both profiles and levels have a hierarchical relationship, and the syntax supported by a higher profile or level must also support all the syntactical elements of the lower profiles or levels. The level deals with the picture resolutions such as the number of pixels per line, lines per frame, frame per seconds (fps) and bits per second or the bit rate (e.g. Mb/s).

Table 1.1: All profiles in MPEG-2 video standard.

| Abbr. | Name | Frames | YUV | Streams | Comments or Applications |
|---|---|---|---|---|---|

| | | | | | |
|---|---|---|---|---|---|
| SP | Simple Profile | I, P | 4:2:0 | 1 | No interlacing |
| MP | Main Profile | I, P, B | 4:2:0 | 1 | Broadcast TV |
| 422P | 4:2:2 Profile | I, P, B | 4:2:2 | 1 | studio post-production etc. |
| SNR | SNR Profile | I, P, B | 4:2:0 | 1-2 | SNR: Signal to Noise Ratio |
| SP | Spatial Profile | I, P, B | 4:2:0 | 1-3 | Low, normal and high quality decoding |
| HP | High Profile | I, P, B | 4:2:2 | 1-3 | |

Table 1.2: All levels defined in MPEG-2.

| Abbr. | Name | Pixel/Line | Lines | Frame Rate (fps) | Bit Rate (Mb/s) |
|---|---|---|---|---|---|
| LL | Low Level | 352 | 288 | 30 | 4 |
| ML | Main Level | 720 | 576 | 30 | 15 |
| H-14 | High 1440 | 1440 | 1152 | 30 | 60 |
| HL | High Level | 1920 | 1152 | 30 | 60 |

### 1.1.2 H.264/AVC

H.264, MPEG-4 Part 10, or AVC, is a digital video standard which is noted for achieving very high compression ratio. It was standardized by the ITU-T Video Coding Experts Group (VCEG) together with the aforementioned ISO/IEC MPEG as the product of a collective partnership effort known as the Joint Video Team (JVT). The ITU-T H.264 standard and the ISO/IEC MPEG-4 Part 10 standard (formally, advanced video coding, AVC) are technically identical. The intent of the H.264/AVC project was to create a standard that would be capable of providing good video quality at bit rates that are substantially lower than what previous standards would need (e.g., relative to MPEG-2, H.263, or MPEG-4 Part 2), and to do so without so much of an increase in complexity. Moreover, an additional goal was to allow the standard to be applied to a very wide variety of networks and systems such as mobile broadcasting, DVD storage, RTP[1]/IP packet networks, and ITU-T multimedia telephony. For example, Table 1.3 demonstrates the existing applications of H.264/AVC. It experienced widespread adoption within a few years of the completion of the standard and is employed widely in applications ranging from television broadcast to video for mobile devices.

Table 1.3: Existing applications for H.264/AVC.

| **Applications** | **Naming** | **Organization or Country** | **CODEC** |
|---|---|---|---|
| DVD | HD-DVD | DVD Forum | MPEG-2, H.264/AVC, VC-1 |
| | Blu-ray Disc | Blu-ray Disc Association | |
| Broadcasting TV | DVB | Europe | H.264/AVC |
| | ATSC | U.S. | H.264/AVC, VC-1 |
| Terrestrial Broadcasting | Prime Minister | France | H.264/AVC |
| | DMB | Korea | |
| | ISDB-T | Japan | |
| Satellite TV | Direct TV | U.S. | H.264/AVC |
| | Euro1080 | Europe | |
| | Premiere | Germany | |
| | Sky Italia | Italy | |
| Telephony | 3GPP[2] | | H.264/AVC |
| Networking | RTP/IP | | H.264/AVC |
| | ISMA | Internet Streaming Media Alliance | |

1 RTP: Real-time Transport Protocol

2 3GPP: The 3rd Generation Partnership Project

| | | | |
|---|---|---|---|
| Video conferencing | | ITU-R BT.1687, 1737 | H.264/AVC |

As mentioned in MPEG-2, H.264/AVC also consists of many "profiles" and "levels" in Table 1.4 and Table 1.5 which are defined as a subset of technologies within a standard usually created for specific applications. The different profiles include or exclude different sets of algorithmic features and coding tools. The inclusion or exclusion of these tools causes an increase or decrease in complexity as well as compression efficiency. For example, the Baseline profile includes the smallest set of compression tools, resulting in the least processing intensive video decoder. It was intended for low delay applications (i.e. w/o B slices) such as video conferencing and video on mobile hand-held devices. The Extended profile can also be used for wireless mobile devices, but it was created primarily for streaming media applications such as those found on today's Internet capable PCs. Moreover, the Main profile has the most compression tools and efficiency gains while High profile targets at high-definition TV applications such as HD-DVD and Blu-ray Disc. The related coding tools adopted by each profile does not mentioned here but can be found in [12] from a high-level perspective.

Table 1.4: All profiles and coding tools in H.264/AVC.

| **Coding Tools** | **B**[3] | **E**[4] | **M**[5] | **High Profile**[6] | | | |
|---|---|---|---|---|---|---|---|
| | | | | **$H_1$** | **$H_2$** | **$H_3$** | **$H_4$** |
| I and P Slices | Y | Y | Y | Y | Y | Y | Y |

[3] Baseline Profile

[4] Extended Profile

[5] Main Profile

[6] $H_1$: High Profile; $H_2$: High10 Profile; $H_3$: High 4:2:2 Profile; $H_4$: High 4:4:4 Profile.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| B Slices | N | Y | Y | Y | Y | Y | Y |
| SI and SP Slices | N | Y | N | N | N | N | N |
| Multiple Reference Frames | Y | Y | Y | Y | Y | Y | Y |
| In-Loop Deblocking Filter | Y | Y | Y | Y | Y | Y | Y |
| CAVLC Entropy Coding | Y | Y | Y | Y | Y | Y | Y |
| CABAC Entropy Coding | N | N | Y | Y | Y | Y | Y |
| Flexible Macroblock Ordering (FMO) | Y | Y | N | N | N | N | N |
| Arbitrary Slice Ordering (ASO) | Y | Y | N | N | N | N | N |
| Redundant Slices (RS) | Y | Y | N | N | N | N | N |
| Data Partitioning | N | Y | N | N | N | N | N |
| Interlaced Coding (PicAFF, MBAFF) | N | Y | Y | Y | Y | Y | Y |
| 4:2:0 Chroma Format | Y | Y | Y | Y | Y | Y | Y |
| Monochrome Video Format (4:0:0) | N | N | N | Y | Y | Y | Y |
| 4:2:2 Chroma Format | N | N | N | N | N | Y | Y |
| 4:4:4 Chroma Format | N | N | N | N | N | N | Y |
| 8 Bit Sample Depth | Y | Y | Y | Y | Y | Y | Y |
| 9 and 10 Bit Sample Depth | N | N | N | N | Y | Y | Y |
| 11 and 12 Bit Sample Depth | N | N | N | N | N | N | Y |
| 8x8 vs. 4x4 Transform Adaptivity | N | N | N | Y | Y | Y | Y |
| Quantization Scaling Matrices | N | N | N | Y | Y | Y | Y |
| Separate Cb and Cr QP control | N | N | N | Y | Y | Y | Y |
| Residual Color Transform | N | N | N | N | N | N | Y |
| Predictive Lossless Coding | N | N | N | N | N | N | Y |

Table 1.5: All levels defined in H.264/AVC.

| Level # | Max. MB/s | Max. frame size (MBs) | Max bit rate for Baseline, Extended and Main Profile | Max bit rate for High Profile | Max bit rate for High 10 Profile | Max video bit rate for High 4:2:2 and High 4:4:4 Profile |
|---|---|---|---|---|---|---|
| 1 | 1485 | 99 | 64kb/s | 80kb/s | 192kb/s | 256kb/s |
| 1.1 | 3000 | 396 | 192kb/s | 240kb/s | 576kb/s | 768kb/s |
| 1.2 | 6000 | 396 | 384kb/s | 480kb/s | 1152kb/s | 1536kb/s |
| 1.3 | 11880 | 396 | 768kb/s | 960kb/s | 2304kb/s | 3072kb/s |
| 2 | 11880 | 396 | 2Mb/s | 2.5Mb/s | 6Mb/s | 8Mb/s |
| 2.1 | 19800 | 792 | 4Mb/s | 5Mb/s | 12Mb/s | 16Mb/s |
| 2.2 | 20250 | 1620 | 4Mb/s | 5Mb/s | 12Mb/s | 16Mb/s |
| 3 | 40500 | 1620 | 10Mb/s | 12.5Mb/s | 30Mb/s | 40Mb/s |
| 3.1 | 108k | 3600 | 14Mb/s | 17.5Mb/s | 42Mb/s | 56Mb/s |
| 3.2 | 216k | 5120 | 20Mb/s | 25Mb/s | 60Mb/s | 80Mb/s |
| 4 | 245760 | 8192 | 20Mb/s | 25Mb/s | 60Mb/s | 80Mb/s |
| 4.1 | 245760 | 8192 | 50Mb/s | 62.5Mb/s | 150Mb/s | 200Mb/s |
| 4.2 | 522240 | 8704 | 50Mb/s | 62.5Mb/s | 150Mb/s | 200Mb/s |
| 5 | 589824 | 22080 | 135Mb/s | 168.75Mb/s | 405Mb/s | 540Mb/s |
| 5.1 | 983040 | 36864 | 240Mb/s | 300Mb/s | 720Mb/s | 960Mb/s |

Based on the illustration of different profiles and levels, we can understand which profile@level is the best candidate for targeted mobile applications. Moreover, this dissertation highlights three design issues to improve the performance when transmitting video over mobile environments. These issues include integration cost, power, and channel impairments, and will be thoroughly discussed in the follow-up Chapters. First of all, the integration and power issues are presented on SP@ML (simple profile at main level) of

MPEG-2 and BL@L4 (baseline profile at level 4) of H.264/AVC. These targeted profiles take advantage of low coding delay since they use no B-frames and hence no backward or interpolated prediction. In addition, the requirements for the frame memory in the simple and baseline profile are smaller than those in the main profile of MPEG-2 and H.264/AVC standards. On the other hand, the proposal supports 1080HD of maximum video resolution and focus on a wide range of decoding resolution. Therefore, aforementioned profiles and levels are suitable for low-delay and low-memory requirements such as mobile phones, video conferencing etc. Second, video over mobile environment is considered in this dissertation as well. To combat the transmission errors in video streams, error resilient coding tools should be supported for a robust transmission. For instance, data partitioning tools can enhance the error-robustness by separating more important data (such as macroblock types and MV values) from less important data (such as inter residual transform coefficient values). However, these coding tools enable only on extended profile of H.264/AVC. Hence, the developed video decoder should support associated coding tools for improving error-robustness and enhancing visual quality to end-users.

### 1.1.3 Motivation of Implementation Methods

The design for realizing multimedia systems on a chip can be roughly partitioned into processor-based and hardwired-based VLSI implementation methods. Several video codecs are optimized for general-purpose processors [13]–[15] (such as ARM, Intel, and Pentium) or DSP processors [16] that have powerful processing capabilities, large memory capacity, special media instruction sets and wide buses to facilitate the reduction of cost and design period. However, for the portable or mobile devices, the design of the optimized video modules is constrained by low computational power and small memory spaces. Hence, processor-based implementation is hard to achieve cost/power efficiency under a specified application field. Moreover, according to the instruction profiling with high-resolution 1080HD video playback via an iprof [18], a new video processor, such as H.264/AVC, requires 83 giga-instructions per second (GIPS) computation and 70 giga-bytes per second (GBPS) memory accesses [19]. Those computational loads are extensive and far beyond the capability of modern general purpose processors. On the other hand, hardwired-based implementation is used for a special-purpose multimedia system to exploit the maximal parallelism and achieve the best performance simultaneously. The optimization of the dedicated hardware design usually fits the target applications. In order to achieve low cost/power requirements, hardwired-based solutions are more suitable than processor-based ones to the design of mobile multimedia systems, and it motivates us to focus on hardwired ASIC implementation in this dissertation.

## *1.2 Organization and Contribution*

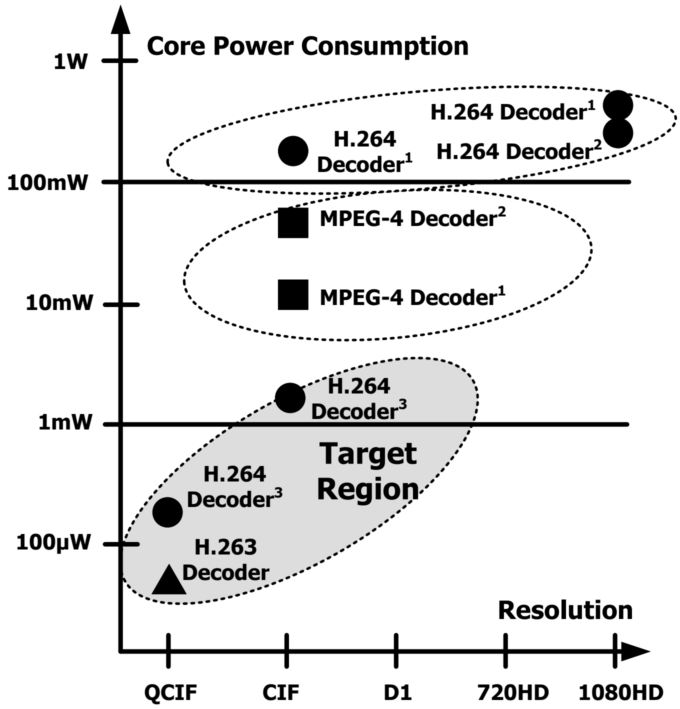
The organization and contribution of this dissertation are demonstrated as follows. We first consider low-cost, low-power and error-robust issues in the developed video decoder. Then, those issues are resolved in algorithmic and architectural levels and further involved in a test chip. Moreover, ASIC design flow, measured results, and detailed comparisons are given as well. Finally, a conclusion has been made separately for highlighting our contributions. A future work has been depicted for further directions to related learners.

### 1.2.1 Low-Cost Design Issue

Considering the standard interoperability for multi-standard applications, this issue addressed integration cost between MPEG-2 and H.264/AVC video standards. There is a great challenge of cost-efficiency for combining MPEG-2 with H.264/AVC due to the standard incompatibility. We first give a similarity analysis and concentrate on the key modules featuring a great diversity between MPEG-2 and H.264/AVC. Those modules include IDCT, entropy decoder, and deblocking filter etc. They have been resolved from algorithmic and architectural levels in Chapter 2. First, we employ a recursive algorithm to cope with the difference of IDCT transformation size, and translate multiplications into a series of shifts and additions to improve hardware utilization. Second, we extract the codeword redundancy in entropy decoders by a table-merging method so as to reduce the table size as well as implementation cost. Third, we develop a joint post-loop and in-loop deblocking filter to meet the different standard requirements. Specifically, post-loop filters can be applied to prevalent MPEG standards and in-loop filters are standardized by H.264/AVC. Both filters can share the data paths and filtering procedures to diminish the integration cost. Finally, this proposal provides a mix of MPEG-2 and H.264/AVC video engine and achieves 20% of cost reduction as compared to a separate design.

### 1.2.2 Low-Power Design Issue

Markets for mobile electronics equipment are currently growing rapidly. Multimedia processing will be an essential function in such mobile-equipment applications. The key technologies to success in these mobile multimedia applications are low power dissipation and high cost effectiveness. Cost effectiveness has been discussed in previous issue. Currently, we concentrate on power issue and give a power distribution of video processing core in existing designs for following illustration. Figure 1.2 exhibits the core power distribution of different types of video decoders according to the decoded video resolution. Due to a great diversity of implementation methods such as process and functionality, this figure just gives a brief overview of existing power magnitude in the video decoding core and excludes the I/O and frame memory power dissipation for simplicity and fairness. All decoding power can be roughly partitioned into three groups: 0.1~1W, 1~100mW, and sub-mW. Several low-power H.264 [22][42] and MPEG-4 [20][21] video decoders are realized to date. However, these levels of power consumption are still not applicable when multimedia capabilities are offered in portable devices. Our design target has been drawn in the gray region. Although Bolcioni *et al.* [17] from *STM Microelectronics* have first developed a sub-mW H.263 video decoder, the power reduction is achieved by exploiting parallel IDCT architectures and skipping zero coefficients for highly quantized video streams. But these low-power techniques may not be properly applied to the newest H.264/AVC due to algorithmic divergences. To this end, we [23] propose three low-power techniques and thereby develop a sub-mW H.264/AVC video decoder. First, a domain-pipelined scalability (DPS) technique is used to optimize the pipelined structure according to the number of processing cycles. Second, three-level memory hierarchy is implemented via a line-pixel-lookahead (LPL) scheme to improve the external bandwidth and reduce the internal memory size, leading to 51% of memory power reduction compared to a conventional design. Third, low-power motion compensation and deblocking filter are designed to reduce the operating frequency without degrading system performance. To summarize, a single-chip MPEG-2 SP@ML and H.264/AVC BL@L4 video decoder is fabricated in a 0.18μm 1P6M CMOS technology with an area of 15.21mm$^2$. This chip contains 19.2kb and 3.55kb of embedded SRAM for storing neighboring pixels and control tags, and adopts two 4MB SDRAMs for further system integration. It operates at a power-level that is about one order of magnitude less than comparable decoders. Furthermore, this low-power design also reveals its strong suitability for mobile electronic equipment where low power requirements are essential.




| Decoder | Process | cf. |
|---|---|---|
| H.264[1] | 0.13μm CMOS | [42] |
| H.264[2] | 0.18μm@1.8V | [22] |
| H.264[3] | 0.18μm@1.8V | [23] |
| MPEG-4[1] | 0.18μm@1V | [21] |
| MPEG-4[2] | 0.3μm CMOS | [20] |
| H.263 | 0.35μm@1V | [17] |

Figure 1.2: Core power dissipation in different types of video decoders.

### 1.2.3 Error-Robust Design Issue

In 2000, Y. Wang *et al.* listed typical bit rates of video data, transmitted packet sizes and error characteristics of practical networks in Table 1.6 [72]. It apparently demonstrated that video transmission over noisy channels is a very challenging task due to the very harsh and time-varying channel conditions. For all cases it is assumed that the application environment does not regularly allow re-transmission of lost or corrupted video packet because of real-time constraints and/or broadcasting environments without backward notification. In this table, although cable/satellite TV or video conferencing over ISDN and ATM provide error-free video transmission channel, there are several types of applications which introduce more or less different error rates. Hence, to combat those transmission errors and improve the subjective visual quality, an error-robust video decoding design has been demanded and will be discussed in Chapter 4. Recently, there has been a lot of interest in the use of soft computing on variable length codes (VLCs) [63]–[71]. In this work, we first develop a soft decoding on context-adaptive VLC (CAVLC) decoder to improve the error detection capability. Moreover, this improved soft CAVLC decoder also facilitates the error concealment due to the correctly or early detected information of corrupted regions. In addition to the improved error detection on soft CAVLC decoder, a joint deblocking filter and error concealment is presented to reduce the implementation cost. On the other hand, a frame re-compression is applied to compress the decoded pixel data and reduce the external memory capacity. It also features error-robustness by skipping the corrupted data when error detection notices the frame re-compression module of error occurrence. Overall, this dissertation just highlights the motivation and brief innovations, and there is still a lot of room for further improving the error-robustness of video decoders.

Table 1.6: Error characteristics in different applications [72].

| Applications | Bit rate for video | Packet size | Error types |
|---|---|---|---|
| ISDN[7] video phone | 64~384 kb/s | N/A | Error free |
| PSTN video phone | 20 kb/s | 100 bytes | Very few bit error and packet loss |
| Mobile video phone | 10~300 kb/s | 100 bytes | BER=$10^{-3}$ ~ $10^{-5}$ |
| Videophone over Packet Network H.323 | 10~1000 kb/s | <=1.5kbytes | BER=0, 0~30% packet loss |
| Cable/Satellite TV | 6~12 Mb/s | N/A | Almost error free |
| Video Conferencing over ATM H.310, H.321 | 1~12 Mb/s | 53 bytes (ATM cell) | Almost error free |

[7] ISDN: Integrated Services Digital Networks