Special Article on Mobile Multimedia Signal Processing Technologies
Video Coding Technology

This article reviews the fundamental technologies of the MPEG-4 video coding standard, which plays a decisive role in IMT-2000 multimedia communications services. It also covers its relationship with prior video coding standards, such as H.261, MPEG-1, MPEG-2 and H.263.

Hiroyuki Yamaguchi, Minoru Etoh and Satoru Adachi

1. Introduction

Multimedia services, including videophones and video distribution, over IMT-2000 mobile networks exploit MPEG-4 visual for video coding. MPEG-4 visual is positioned as a compilation of the video coding technologies developed to date.

This article reviews the fundamental technologies and properties of various video coding technologies that preceded MPEG-4 visual.

2. Fundamental Technologies of Video Coding

Video signals normally add up to a huge amount of information, e.g., around 100 Mbit/s. Various highly efficient video coding technologies have been invented to exploit video signal properties. Three fundamental technologies are commonly used for video coding. These are motion-compensated interframe prediction, Discrete Cosine Transform (DCT) and Variable Length Coding (VLC) [1]-[3].

(1) Motion-compensated Interframe Prediction

Motion-compensated interframe prediction refers to interframe coding that identifies the magnitude and direction of the motion of certain parts in the image by obtaining the difference from frame to frame, instead of coding all of the frames (Figure 1). As the magnitude and direction of motion (motion vector) vary throughout the picture, the frame is divided into small blocks (about 16×16 pixels each) called macroblocks. Then, a motion vector is obtained for each macroblock, in which motion is regarded as being the difference with the previous frame. The difference is called the prediction error, and is subjected to (2) Discrete Cosine

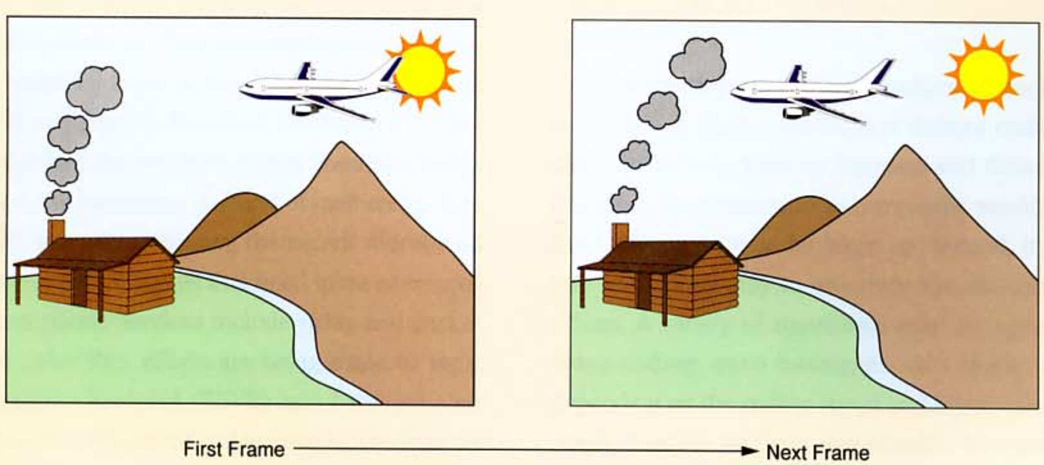

Figure 1 Concept of Motion-compensated Interframe Prediction

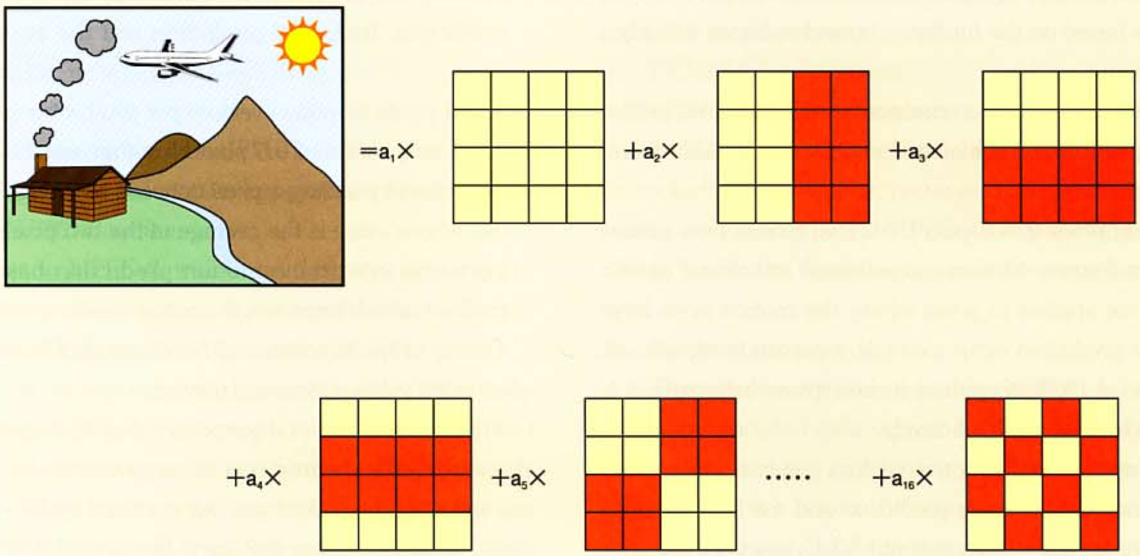

Figure 2 Decomposition of Image into Frequency Components

Transform (DCT), as described below.

(2) Discrete Cosine Transform (DCT)

DCT is a technique for representing each video frame as a weighted sum of simple (low-frequency) image components and complex (high-frequency) image components (Figure 2). It is known that visual information tends to concentrate on low-frequency components, playing a substantial role in visual perception. DCT aims to extract the important frequency components and compress the visual information. DCT is widely used, as it is able to effectively transform images to the spatial-frequency domain.

In practice, the frame is divided into blocks of about 8×8 pixels, in order to apply DCT to each block. "ai" in Figure 2 is referred to as the DCT coefficient, which is to be quantized and rounded off to a representative figure. This process is followed by Variable Length Coding (VLC).

(3) Variable Length Coding (VLC)

VLC exploits the uneven value distribution in the input signal in order to compress information. It assigns shorter codes to frequent signals and longer codes to less frequent signals.

When high-frequency coefficients are quantized and rounded off, as described in the DCT section, many of them will end up as "zero" coefficients. As a result, the occurrence of "End of Block" (EOB) and "value L after a certain number of zeros" will increase. Shorter codes can be assigned to combinations of the number of zeros (zero-run) and the value of L (Level) that occur frequently, in order to compress the information. This technique, which assigns a code to each pair, is called two-dimensional VLC.

3. Positioning of Video Coding Standards

The major video coding international standards are H.261, MPEG-1, MPEG-2, H.263 and MPEG-4. Figure 3 illustrates the positioning of each standard. The following sections explain how the previously described fundamental technologies are used to improve the compression efficiency of each standard and how their functions differ.

(1) H.261 Video Coding

This is practically the first international video coding standard for videotelephony and videoconferencing over Integrated Services Digital Network (ISDN). H.261 was recommended by the International Telecommunication Union-

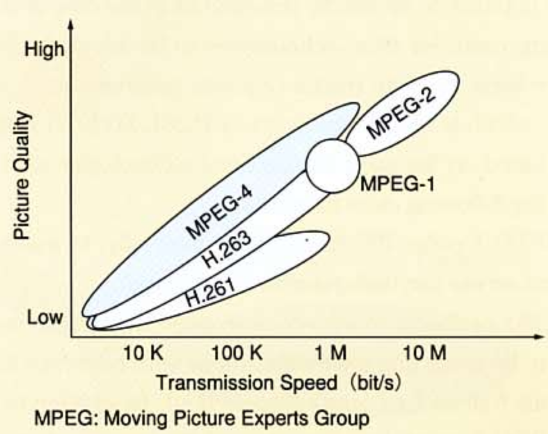

Figure 3 Relationship between MPEG-4 Visual and Other Standards

Telecommunication Standardization Sector (ITU-T) in 1990 [4], and is based on the fundamental technologies described in the previous section.

- H.261 performs motion-compensated interframe prediction; it predicts pixcel-wise motion vectors for 16 × 16 macroblocks.
- H.261 applies 8 × 8-pixel DCT to prediction errors between frames. Motion-compensated interframe prediction is not applied to areas where the motion is so large that the prediction error exceeds a certain level; instead, 8 × 8-pixel DCT is applied to in-a-frame information to achieve better coding efficiency.
- VLC is applied to the motion vectors generated by motion-compensated interframe prediction and the DCT processing results. Two-dimensional VLC is executed with respect to the DCT processing results.

H.261 was originally designed to coexist with the cameras and monitors of standard television. As the format of TV signals (the number of frames and the number of scanning lines) depends on the region, a proposal was made to convert the signals into a "neutral," common format before coding them, in order to make international communications possible. This is called Common Intermediate Format (CIF), and defined as "352 (horizontal) × 288 (vertical) pixels, up to 30 frames per second, and non-interlaced." Quarter CIF (QCIF), which accounts for a quarter of the resolution of the full CIF, was defined at the same time and is used for video coding.

(2) MPEG-1 and MPEG-2 Video Coding

MPEG-1 was standardized by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) in 1993, for CD-ROMs and other storage media [5]. MPEG-1 uses a bitrate of around 1.5 Mbit/s. For storage media, the encoding process does not have to be strictly real-time as in the case of H.261, leaving room for new technologies to be adopted. On the other hand, storage media requires random search functions, which is beyond the scope of H.261. MPEG-1 is therefore based on the same fundamental technologies as H.261, with the following extra functions.

- MPEG-1 codes the entire frame intervally, to make random access playback possible (Intraframe).
- H.261 performs motion-compensated interframe prediction, by predicting the motion vector with reference to previous frames (i.e., forward prediction). In addition to this, MPEG-1 can take advantage of the merits of storage media and predict motion from future frames (i.e., backward prediction). The compression rate is enhanced by working out the minimum prediction error from forward prediction, backward prediction and the average of the two.
- H.261 predicts motion vectors per pixel, whereas MPEG-1 predicts in units of 0.5 pixel. In other words, a frame is generated by adding a pixel between two neighboring pixels, whose value is the average of the two pixels. MPEG-1 performs interframe motion prediction based on this method, which improves the compression rate.

Owing to these additional functions, MPEG-1 is widely used in PC video coders and players.

MPEG-2 is a general-purpose video coding technology that meets the requirements of communications, broadcasting and storage applications. Standardized by ISO and IEC in 1996, MPEG-2 shares the same text with ITU-T H.262 [6] and is used for video coding at 3 bit/s-20 Mbit/s for High Definition Television (HDTV) and DVD. In addition to the fundamental technologies used in MPEG-1, it is equipped with the following extra functions.

- MPEG-2 can efficiently encode interlaced pictures in standard TV format.
- MPEG-2 can adjust the picture size and quality by extracting part of the encoded data as required (called Spatial Scalability and Signal-to-Noise (SNR) Scalability, respectively).

These additional functions, which were aimed at tackling various applications, give rise to the need to ensure compatibility between encoded data. MPEG-2 solves the compatibility problem by categorizing functional differences and processing complexity through the introduction of two new concepts: profile and level. The same concepts are adopted by MPEG-4.

(3) H.263 Video Coding

H.263 was standardized by ITU-T in 1996 as a technology for ultra-low bitrate video coding in analog telephone networks [7]. It assumes the use of 28.8 kbit/s modems and includes some new technologies developed for MPEG-1, which were not available at the time that H.261 was standardized. The basic functionality of H.263 (H.263 Baseline) mandates motion-compensated interframe prediction in units of 0.5 pixel. H.263 Baseline also has a three-dimensional feature, adding EOB to the two dimensions (run, level) in order to boost coding efficiency. New options include motion-compensated interframe prediction in 8 × 8 blocks and processing to mitigate block-shaped distortions in the picture.

Some ISDN videotelephony/videoconferencing equipment optionally adopt H.263 for video coding to achieve the above additional functions.

4. MPEG-4 Visual Coding

MPEG-4 visual is based on ITU-T H.263, with various enhancements made to increase error robustness. It is compatible with H.263 Baseline.

In contrast to MPEG-2, which was designed primarily for handling videos on computers, digital broadcasting and high-speed communications, MPEG-4 was standardized with a strong focus on the communications sector, especially mobile communications. As a result, it was finalized by ISO and IEC as an extremely versatile video coding technology in 1999 [8]. It is regarded as the key technology not only for videotelephony over the IMT-2000, but also for video-based multimedia services including video mail and distribution (Figure 4).

(1) Profile and Level

MPEG-4 categorizes functionality by profile and required processing power by Level — in the same manner as MPEG-2 — in order to ensure the shared use of encoded data, including interoperability. By definition, Profiles include Simple, Core, Main and Simple Scalable, out of which the Simple Profile serves as the common function. 8×8 motion-compensated interframe prediction, which was optional for H.263, is defined as a Simple Profile.

In the Simple Profile, Level 1 is assigned to QCIF images and Level 2 to CIF images.

Meanwhile, the Core and Main Profiles define any part of the video as an object so that its picture quality can be improved, or be incorporated into other encoded data. MPEG-4 also supports advanced profiles that synthesize computer graphic images.

(2) IMT-2000 Specifications

3G-324M specifies visualphone over the IMT-2000, which is described in detail in the article titled audiovisual terminal technology in this volume, requires H.263 baseline for video coding, and recommends MPEG-4 Simple Profile, Level 1. Simple Profile has the following tools to ensure error robustness.

① Resynchronization

This involves the insertion of a resynchronization code into the encoded data subject to VLC, in order to divide the picture at the right place and localize transmission errors. As the header information is inserted after the resynchronization code to specify the coding parameters, data affected by transmission errors can be promptly recovered. The insertion interval of the resynchronization code can be optimized with reference to the overhead of the header information, the subject of the recording and the properties of the transmission channel.

② Data partitioning

This involves the insertion of a synchronization code at the borderline between different data types in the encoded data. It enables error concealment. For example, if there are any bit errors in the DCT coefficient, the insertion of a synchronization code between the motion vector and the DCT coefficient will ensure the accurate transmission of the motion vector, enabling error concealment in a more natural manner.

③ Reversible Variable Length Code (RVLC)

As shown in Figure 5, Reversible Variable Length Code

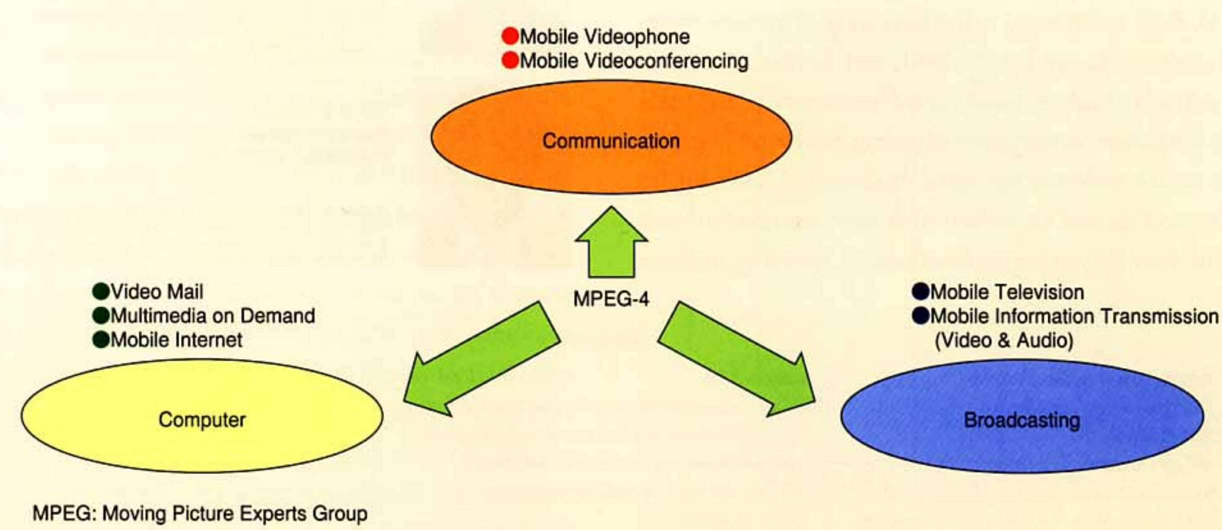

Figure 4 Scope of MPEG-4 Visual

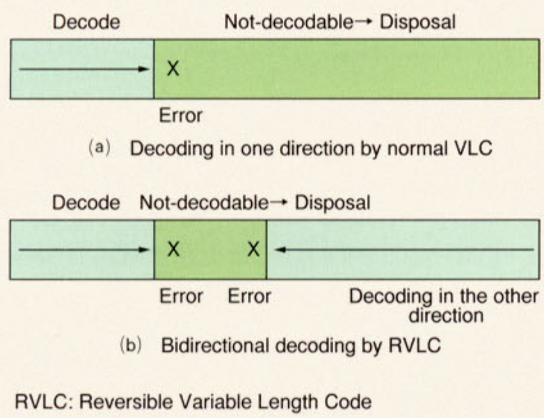

Figure 5 Example of RVLC Decoding

(RVLC) can be decoded from the opposite direction. By applying this to the DCT coefficient, any data other than macroblocks with bit errors can be decoded.

④ Adaptive Intra Refresh (AIR)

This is a technique to execute frequent intra-frame coding to an area to minimize error propagation.

With the use of MPEG-4 Simple Profile, Level 1, as described above, an extremely simple coder decoder can be created, suitable for mobile communications.

5. Conclusion

Seeing is believing. This saying describes the enormous potential of video-based multimedia services, which entail a massive volume of video data. The progress of video coding technology has been marked by the finalization of MPEG-4 visual, while wireless access technology has seen the development of Wideband Code Division Multiple Access (W-CDMA). Both contributed to the launching of wireless video-based multimedia services in 2001, and further efforts are expected to be made in boosting the performance and minimizing the power consumption of coding hardware.

This article reviewed the three fundamental technologies in video coding and described how they are used in each standard, how they enhance functionality, how they improve the compression ratio of data, and how they are reflected in the MPEG-4 standard.

It could be said that these three fundamental technologies form the core in this field. However, challenges are continually being made at academic conferences to develop new technologies, which deserve much attention. Further progress can be expected in this area.

References

[1] Hiroshi Yasuda: International Standards of MPEG/Multimedia Coding, Maruzen Co., Ltd. (1994).

[2] Hiroshi Fujiwara: Latest MPEG Textbook, ASCII Corp. (1994)

[3] G.Sullivan and T.Wiegand: "Rate-Distortion Optimization for Video Compression." IEEE Signal Processing Magazine, 15, 6. pp.74-90 (Nov.1998) .

[4] ITU-T Rec. H.261 (1993), Video codec for audiovisual services at p×64 kbit/s.

[5] ISO/IEC 11172-2: 1993, Information technology — Coding for moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s-Part 2: Video.

[6] ITU-T Rec. H.262 (1995) ¦¦ ISO/IEC 13818-2: 1996, Information technology-Generic coding of moving pictures and associated audio information: Video.

[7] ISO/IEC 14496-2: 1999, Information technology — Generic coding of audio-visual objects — Part 2: Visual.

Glossary

CIF: Common Intermediate Format
DCT: Discrete Cosine Transform
EOB: End of Block
HDTV: High Definition Television
IEC: International Electrotechnical Commission
IMT-2000: International Mobile Telecommunications-2000
ISO: International Organization for Standardization
ITU-T: International Telecommunication Union-Telecommunication Standardization Sector
MPEG: Moving Picture Experts Group
QCIF: Quarter CIF
RVLC: Reversible Variable Length Code
SNR: Signal to Noise Ratio
W-CDMA: Wideband Code Division Multiple Access