Chapter 4
The Development and Standardization of Ultra High Definition Video Technology

Tokumichi Murakami

1 Introduction

Video technology has evolved from analog to digital and SD (Standard Definition) to HD (High Definition). However, to provide a visual representation with high quality that satisfies the full range of human visual capabilities it requires further advances in video technology. One important direction is ultra high resolution video. Although UHD (Ultra High Definition) has already been standardized as a video format with spatial resolution 3840x2160 and 7680x4320 in an ITU recommendation (ITU-R BT.1769), actual deployment of UHD services have not yet been realized.

In order to realize UHD video services, the basic technologies that support UHD video, such as high quality camera, display, storage and transmission infrastructure, are indispensable. Presently, these technologies have accomplished remarkable progress, and the video and the visual equipments with 4Kx2K (4K) or 8Kx4K (8K) resolutions exceeding HD are shown at many trade shows and exhibitions. Also, several cameras corresponding to 4K have already been announced, and there are a variety of displays, such as liquid crystal displays (LCD), plasma display panels (PDP) and projectors, which can render 4K video. Moreover, organic electroluminescence (organic EL) equipped with thinness, power saving and high resolution is also promising as a UHD display. Furthermore, the Japan Broadcasting Corporation (NHK) has developed a 33 million pixel camera for a Super Hi-Vision system with 7680×4320 resolution, and is demonstrating a projector and a liquid crystal panel with 8K resolution. Thus, the realization of UHD video service is within reach.

Tokumichi Murakami
Mitsubishi Electric Corporation
Research and Development Center
5-1-1 Ofuna, Kamakura City, Kanagawa, 247-8501 Japan
Tel.: +81-467-41-2801
e-mail: Murakami.Tokumichi@eb.MitsubishiElectric.co.jp

High performance video coding technology is another indispensable element to realize UHD video. At present, the main standards for video coding are MPEG-2 and MPEG-4 AVC (Advanced Video Coding)/H.264 (AVC/H.264). However, the development of a new video coding technology is necessary for UHD video applications since the video must be compressed further to be transmitted within current systems, while still keeping the high quality of the original source as much as possible. In response to such environmental conditions and demands, the standardization activity of a next generation video coding for UHD video is getting underway.

In this chapter, the history and international standardization of video coding technology are described. Then, the fundamental constituent factors of video coding are introduced. Next, the requirements for the video coding technology towards the realization of UHD video are described, and the progress of the supporting UHD video technologies is surveyed. Finally, the challenges toward the technical development of a next generation video coding and the view of future video coding technology are discussed.

2 Progress of Digital Video Technique

Looking back upon the history of video technology, it is evident that video coding is one of the most important elements when considering the progress of digital video technology. During these two decades, MPEG (Moving Picture Experts Group) has occupied a central position in the international standardization of video coding technology. In this section, the results of MPEG standardization and current activities are surveyed.

2.1 History of Video Coding

2.1.1 Before AVC/H.264

Television broadcasting first started as analog in the 1940s and spread generally and widely. In the telecommunications area, video transmission was realized as TV phone service at the beginning, which was sending semi-video in addition to voice through the telephone line. However, it was not practical at that time to allocate wide range of the bandwidth of the communication line for video transmission. Although research and development for efficient video transmission were conducted during the 1950s and 1960s, most of them were video transmission systems made use of analog technology.

In the 1970s, since the digital signal processing began to evolve into practical and more matured technology, the hierarchy of the digital telecommunications network was specified in the communications field. In the beginning of 1980s, the practical development on high efficiency digital video compression had come accelerated. As a result, innovative video coding technology was introduced into the video conference systems for business use, by KDD, NTT and Mitsubishi etc., in Japan. In the middle of 1980s, it became possible to simulate the video coding algorithms more easily with improved workstation capability, and practical research was greatly advanced. As it was at the dawn of a new age of digital communication line based on Integrated Services Digital Network (ISDN), the development of the products of pioneering video conference system was carried out. The CCITT (Consultative Committee for International Telephone and Telegraph), now known as ITU-T (International Telecommunication Union – Telecommunication sector), began to consider and discuss the needs of interconnectivity and interoperability for video transmission assuming TV phone, video conference, remote surveillance, etc. In 1990, CCITT has recommended H.261 for video coding scheme at the transmission rate of px64 Kbit/s (p=1, 2, ...) for the communication of video and audio on ISDN [1]. In H.261, a hybrid coding system using the combination of motion compensated prediction coding and transform coding was adopted, and many of the current video coding systems are derived from the hybrid coding system from H.261. In this period, VTR became more widespread in the home because of its tendency of lower-pricing. Under such a situation, MPEG, which has been a working group under ISO/IEC, started the development of an international standard for video coding that aimed at consumer appliances in 1988.

Since the 1980s, the history of video coding technology has been deeply related to the international standardization. MPEG specifically aimed at the development of audio and video coding methods for CD (Compact Disc) which began to spread rapidly with the advent of digital music. MPEG-1 specified the coding of video with up to about 1.5 Mbit/s; this standardization was completed in 1993 and was adopted for video CD and CD karaoke [2].

Subsequently, MPEG-2 was standardized as MPEG-2/H.262 [3, 4] and aimed at the coding of SDTV and HDTV; this standardization was completed through cooperation between ISO/IEC and ITU-T in 1994. The standardization of MPEG-2 triggered the roll-out of digital broadcasting. Video coding technology provided a means to satisfy constraints on the communication bandwidth and storage capacity for transmitting and storing video, respectively. Until MPEG-2 was standardized, the video was always treated with lower resolution than standard television broadcasting. However, with the advent of MPEG-2, video coding technology was able to realize high quality video services. In 1995, NHK and Mitsubishi had jointly developed an HDTV codec conforming to MPEG-2 specification, and conducted a verification experiment on digital HDTV broadcasting, This became a turning point to accelerate digital TV broadcasting. After then, MPEG-2 began to be adopted as a video coding scheme for digital broadcasting in Japan, Europe and the United States. HDTV digital broadcasting began to be a full-fledged service world-widely in early 2000s, and the spread of LCD displays brought the realistic video experience of 1920 scanning lines to home.

On the other hand, the combination of the Internet and PC had grown greatly as a platform for multimedia services since Mosaic, which was an Internet browser, first released in 1993. In that period, many proprietary coding methods were developed outside the conventional standardization organizations. In spite of this situation, H.263 [5] and MPEG-4 [6] were still used in many applications. H.263 was recommended for the transmission of VGA (Video Graphic Array) video from tens of Kbit/s to several Mbit/s in 1995. MPEG-4 was a successor of H.263 and was completed its standardization in 1999. MPEG-4 was utilized for 3G mobile video phones with 64 kbit/s, portable video players with up to 2-3 Mbit/s, as well as the animation function of digital still camera, etc.

In the 2000s, the development of video coding technology progressed rapidly due to an increase in processing speed of devices, user demand for higher quality and an abundance of video services. AVC/H.264 [7, 8] is the standard which was developed based on the coding techniques examined under the H.26L project in ITU-T/SG16/Q.6 known as VCEG (Video Coding Experts Group). A collaborative team known as the JVT (Joint Video Team) was formed between MPEG and VCEG in 2001. AVC/H.264 which achieved twice as much compression ratio of MPEG-2 was standardized in 2003. The AVC/H.264 standard was then adopted as the coding method for mobile TV broadcasting called One-seg in Japan and for Blu-ray with HD resolution, and continues to extend the scope of its applications.

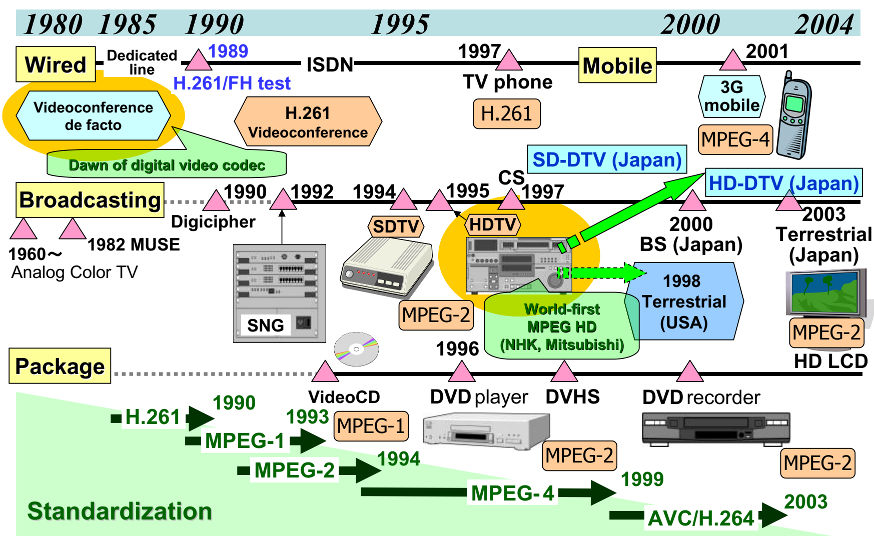


Fig. 1 History of Video Technology (From Visual Communication to Digital Broadcasting)

2.1.2 After AVC/H.264

AVC/H.264 continued to be improved for high quality video after the recommendation in 2003. Baseline Profile including 4x4 transform was standardized in 2003 and High Profile which employs 8x8 transform and individual quantization matrix for HDTV was recommended in 2005. High Profile has been adopted and deployed in home AV equipments including Blu-ray disc players and recorders. In 2007, some additional coding tools to support high quality video coding were added, including the support for coding of video in 4:4:4 format and high definition levels, according to the proposals from Mitsubishi, etc. [9].

In response to this progress in video coding technology, television broadcasting, which is the most familiar video media to the public, has started shifting from SDTV to HDTV in a digital form. Video recorders for home use and small handheld camcorders are also operated with HDTV quality; such devices realize not only a small size but a low price as well.

Thus, while HD video is becoming the norm, the development of UHD video technology such as 4K with 4 times the resolution of HD is progressing steadily. Visual equipments with 4K resolution are now being exhibited at shows. Several cameras corresponding to 4K resolution have already been announced, and LCD, PDP and projectors which can display 4K image can be seen. Moreover, the organic EL equipped with thinness, power saving and high resolution is also promising as a display of UHD video. With respect to practical use, digital cinemas with 4K resolution have been specified and their use for digital cinema including distribution to theaters has already started [10]. Furthermore, NHK is planning for advanced television broadcasting with 8K resolution from the year of 2025, and already has developed a 33 million pixel camera with 8K resolution and a projector with 8192x4320 resolution. Video standardization of 4K and 8K resolutions is being progressed by ITU-R (Radiocommunications sector) and SMPTE (Society of Motion Picture and Television Engineers) which is responsible for production standards used by the cinema and television industries. Next-generation video coding standards, including UHD video as a target, are also in the process of starting in response to these environmental conditions and expectations.

In 2009, MPEG invited the public to submit evidence of new video coding technologies that fulfill the conditions for UHD video, and evaluated the technologies considering the emerging application requirements [11]. As a result of this study, sufficient evidence was obtained and MPEG is now planning a new standardization initiative to meet these goals. The current schedule is to collect proposals in 2010, and to recommend an international standard in 2012-2013. MPEG and VCEG are likely to cooperate on this activity.

Thus, information projected on a screen will be diversified in the future when UHD video technology for 4K and 8K resolutions is realized. For example, we will be able to enjoy realistic and immersive video on a large screen TV that is over 100 inches diagonal, and display web browsers simultaneously with UHD video contents on the screen. We may also use a photogravure TV of A3 size like an album of photographs.

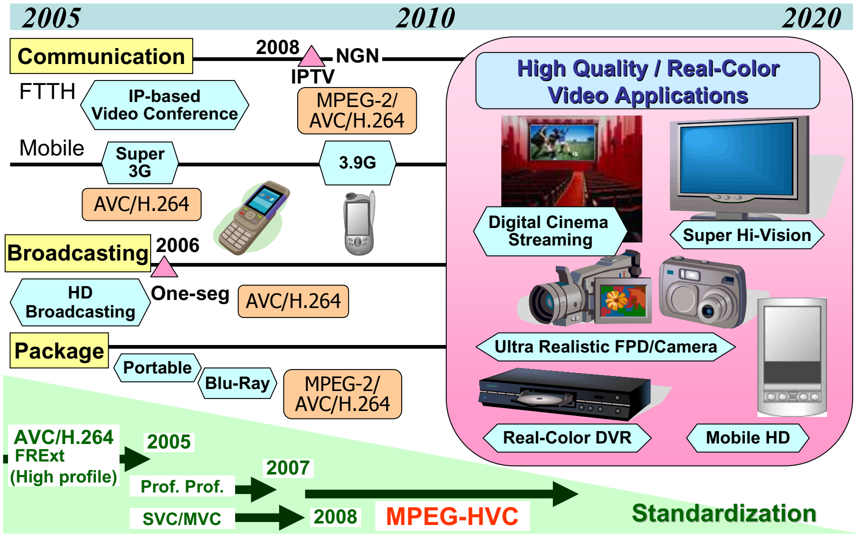


Fig. 2 History of Video technology (Nowadays and Future)

2.2 Technical Standardization for Video Coding

The history of video coding technology has been deeply related to international standardization since the 1980s. Because interoperability is very important for the widespread utilization of video contents, it is important that coded video contents should conform to an international standard. In the following, international standards for video coding focusing on MPEG are surveyed.

2.2.1 International Organizations of Video Coding Standards

Video coding technology has been playing an important role in the progress of video technology and the spread of video contents and applications. Standardization organizations responsible for digital video related technologies and their mutual relationships are shown in Fig. 3.

ISO/IEC and ITU-T are the primary organizations in the world that engage in the international standardization of video coding. MPEG is one of the working groups, formally known as ISO/IEC JTC1/SC29/WG11, which belongs to a Joint committee of ISO (International Standardization Organization), which promulgates industrial and commercial standards, and IEC (International Electro-technical Commission) which treats international standards for all electrical, electronic and related technologies. On the other hand, ITU-T is an international standardization organization for telecommunications which was formerly called CCITT, and it has been coordinating many standards in the field of telecommunications.

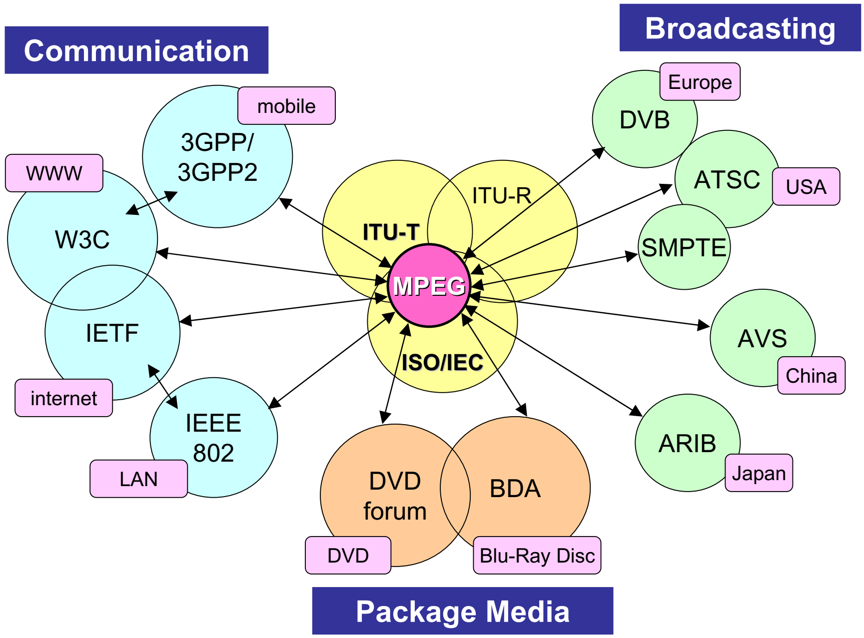


Fig. 3 Standardization Organizations and their Relationships

Among the digital video related standardization organizations, MPEG has been and continues to occupy the central position of video coding technology, and MPEG standards are considered as one of the essential technologies for digital video services. Since its inauguration in 1988, MPEG has standardized MPEG-1, MPEG-2/H.262, MPEG-4 and AVC/H.264, and has been promoting the development and standardization of multimedia technologies including video coding. MPEG standards specifies the technologies to compress video data with compression ratios in the range from 30:1 to 100:1 as well as the technologies for transmission and storage of video and audio contents, and offers open specifications and compatibilities.

Furthermore, MPEG has cooperated on international standardization of video coding with VCEG (Video Coding Experts Group) which is affiliated with ITU-T/SG16. Experts from both MPEG and VCEG committees formed the JVT (Joint Video Team) for the development of AVC/H.264.

2.2.2 Improvement in Compression Ratio by MPEG

The compression ratio of video data has improved through MPEG standardization. For example, AVC/H.264 can perform twice as much compression ratio as MPEG-2. High resolution and multi-channel were attained by the improvement of the compression ratio of video coding by MPEG.

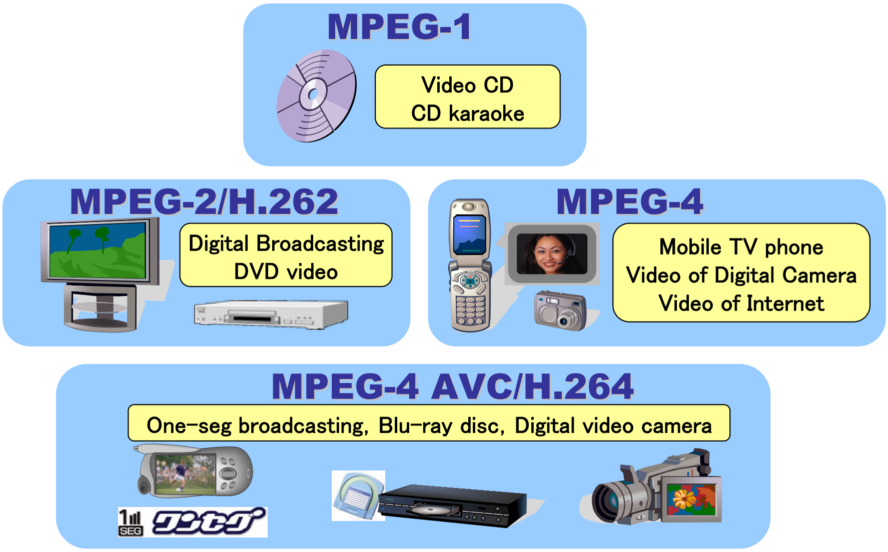


Fig. 4 MPEG and Digital Video Services

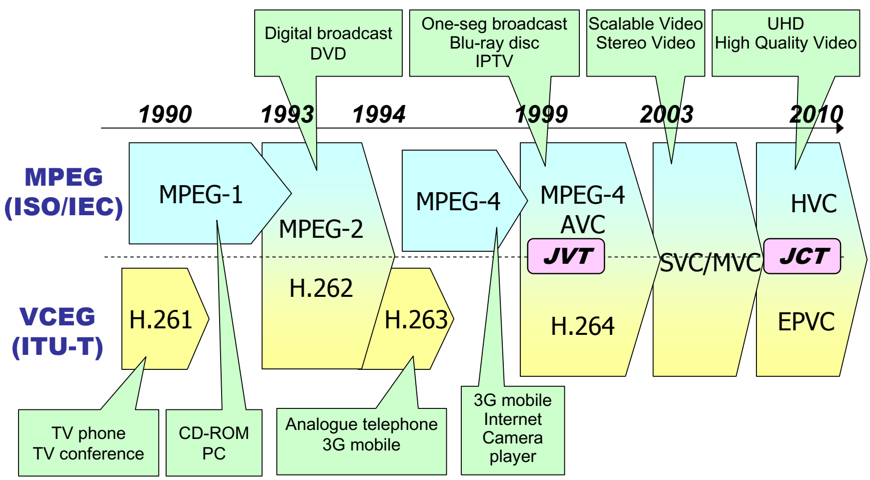


Fig. 5 Cooperative Relationship between MPEG and VCEG

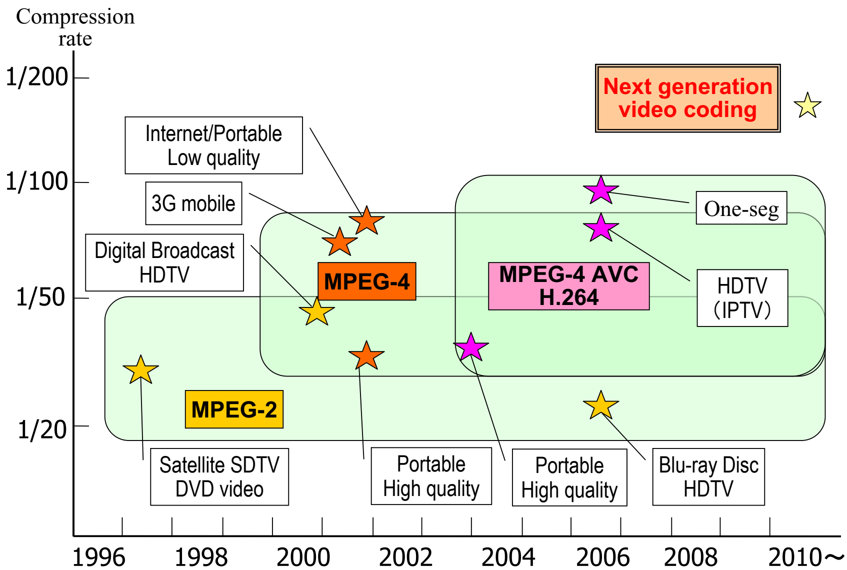


Fig. 6 Improvement in Compression Ratio by MPEG

3 Video Coding Technology

3.1 The Shannon Theory and Video Coding Technology

Video coding aiming at digital information compression began its development from the necessity of transmitting a vast quantity of digital video data through communication line with narrow band. Therefore, it is possible to draw an analogy between the information and telecommunication model of Shannon [12, 13] and the composition of digital video coding transmission system (refer to Fig. 7). In the sending side of a digital video transmission system, the analog video signal acquired from a camera is digitally sampled and quantized. Format conversion is performed using various filters and a sequence of digital images is generated [14]. Then, prediction, transform, quantization and entropy coding are applied to the image sequence to produce a compressed bitstream (Source Coding) [15]. The compressed bitstream then undergoes encryption, multiplexing, error detection/correction, modulation, etc., and is transmitted or recorded according to the characteristic of a transmission line or a recording medium (Channel Coding). On the other hand, in the receiving side, the video signal is reproduced by the inverse operations performed in the sending side and the video is displayed on a screen.

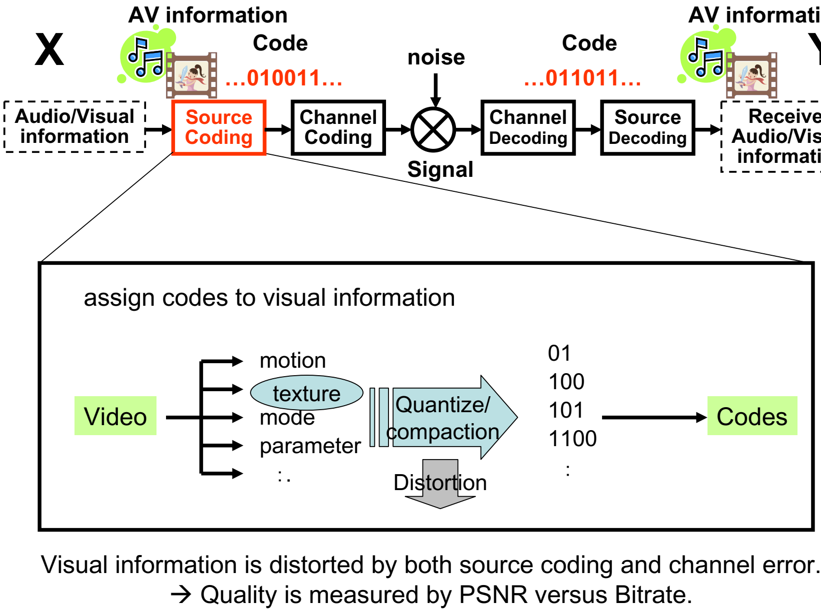


Fig. 7 Video Coding in Shannon's model

3.2 Main Components of Video Coding

Video coding usually consists of four main components including prediction, transform, quantization and entropy coding. Prediction reduces relative redundancy exploiting correlation within a picture or across several pictures. The residual signal that represents the difference between the original and the predicted signal is encoded. Transform is a process for energy compaction of the signal to reduce the correlation of the symbols. In practice, the signal is transformed from a spatial domain to a frequency domain. There are several transforms that have been used in typical image and video coding standards including Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT). Quantization is a technique that reduces the amount of information directly. There are two main methods of quantization including Scalar Quantization and Vector Quantization. Entropy Coding is a reversible coding method based on statistical characterization of the symbols to be encoded. Huffman coding and arithmetic coding are typical examples of entropy coding schemes.

3.2.1 Prediction

A picture has high correlation between neighboring pixels in both spatial and temporal directions. Consequently, the amount of information can be reduced by the combination of the prediction between pixels and the coding of the prediction error (residual signal). The prediction exploiting spatial correlation within a picture is known as Intra prediction, while the prediction exploiting temporal correlation across two or more pictures is known as Inter prediction. A method of further exploiting correlation between frames is to utilize motion prediction, which is referred to as Motion Compensated Prediction. Fig. 8 shows the difference of power between several signals in a typical video coding system.

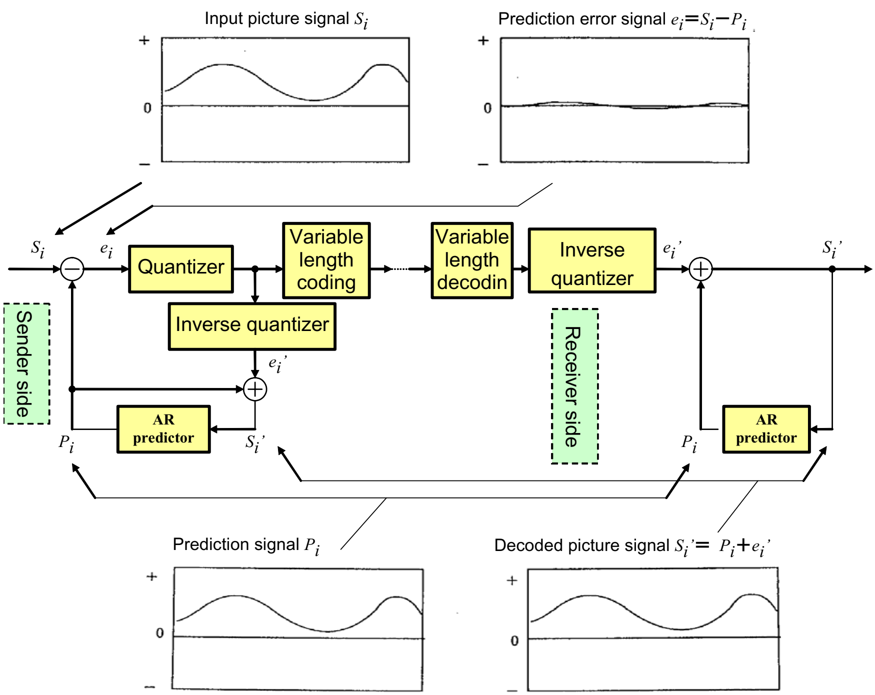


Fig. 8 Predictive Coding Scheme

3.2.1.1 Intra Frame Prediction

Intra Frame Prediction is a prediction technique that uses the neighboring pixels within a frame. Three prediction methods including Previous-sample Prediction, Matrix Prediction and Plane Prediction are shown as examples of Intra Frame Prediction in Fig. 9. Previous-sample Prediction uses neighboring pixels in the horizontal direction as a prediction pixel, Matrix Prediction uses neighboring pixels in both horizontal and vertical directions, and Plane Prediction uses neighboring pixels in horizontal direction and subtracts the pixel values at the same positions on the former line.

3.2.1.2 Motion Compensated Prediction

Motion Compensated Prediction is a technique which creates a prediction image that resembles the current image by linear translation of a block within a reference picture which is already transmitted and decoded. Compression is achieved by coding the difference between the predicted and original pictures. The principle of Motion Compensated Prediction is shown in Fig. 10.

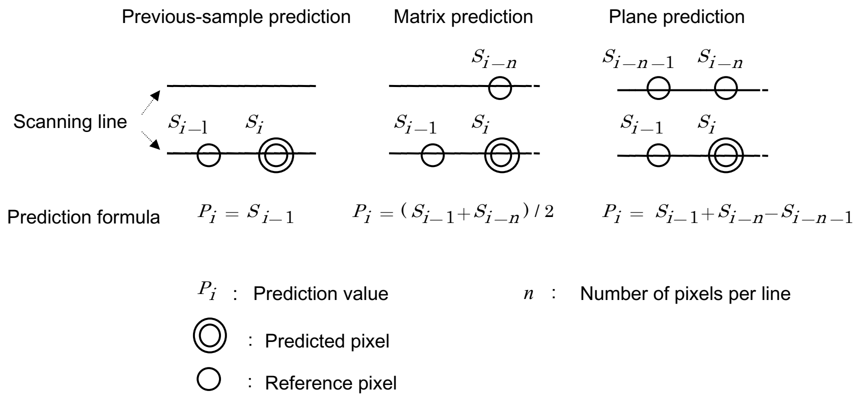


Fig. 9 Examples of Intra Frame Prediction

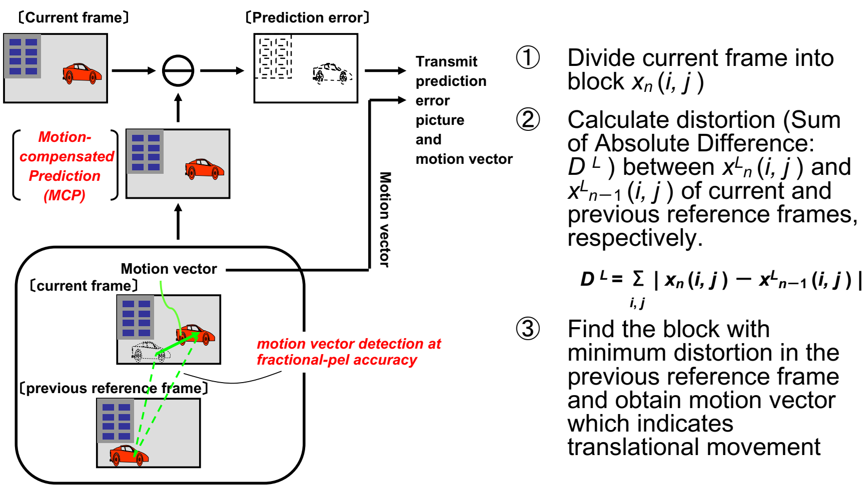
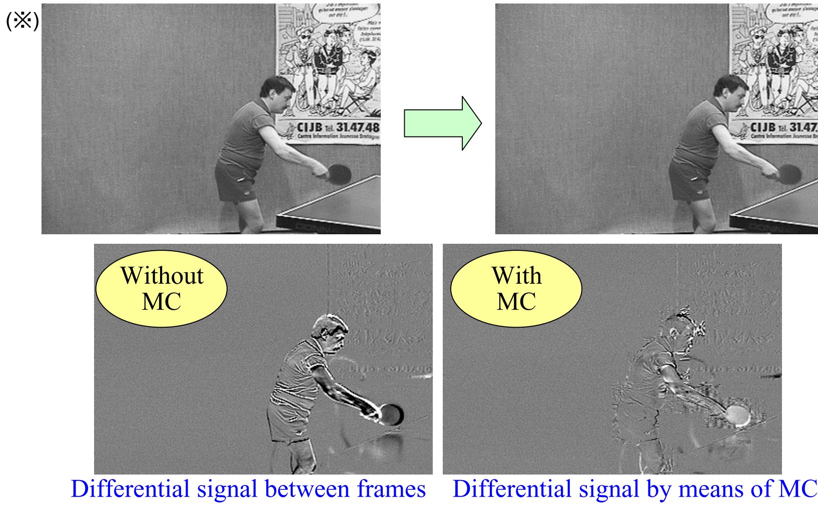


Fig. 10 Principle of Motion Compensated Prediction

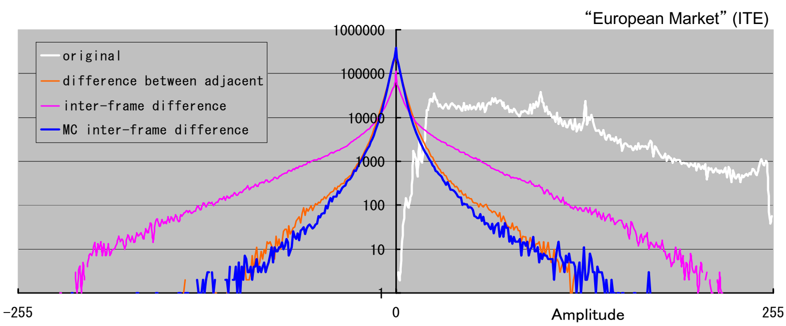
Motion Compensated Prediction can reduce the energy of the residual signal compared with the simple difference between frames. Fig. 11 shows an example that compares the simple difference signal between frames and the Motion Compensated Prediction difference signal. It is clear that the difference signal decreases dramatically when Motion Compensated Prediction is utilized.

Fig. 12 shows an example of the signal characteristics of the original, Intra Frame Prediction, Inter Frame Prediction and Motion Compensated Prediction pictures of a HDTV picture with their entropy and signal power. It is shown that signal power decreases sharply when Motion Compensated Prediction is utilized.



(※) "Table Tennis" (MPEG standard image)

Fig. 11 Effect of Motion Compensated Prediction



	Original signal	Differential signal between adjacent pixels	Differential signal between frames	Motion compensated differential signal between frames
Entropy [bit/pixel]	7.12	4.37	6.05	**4.16**
Signal Power (Variance)	118.22	47.89	10.63	**0.58**

Fig. 12 Characteristics of Picture Signal

3.2.2 Transform

Transform is the method of converting an image signal into another signal domain, and centralizing signal power to specific frequency bands. There exist DCT and DWT for this purpose, which are used in the current picture coding standards.

3.2.2.1 DCT

DCT converts the spatial domain signal into the frequency domain using a window with fixed width for the transformation. Usually, a picture is divided into NxN pixel blocks (N pixels width both horizontal and vertical directions) and the transform is performed for each pixel block. The DCT is expressed as follows,

$$F(u,v)=\frac{2}{N}C(u)C(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}f(x,y)\cos\left[\frac{(2x+1)u\pi}{2N}\right]\cos\left[\frac{(2y+1)v\pi}{2N}\right]$$

where

$$C(u),\ C(v)=\begin{cases}\frac{1}{\sqrt{2}} & (u,\mathrm{v}=0)\\ 1 & (u,\mathrm{v}\neq 0)\end{cases}.$$

On the other hand, the inverse transform (IDCT) reconverts a transformed signal to the spatial domain and is expressed as follows,

$$f(x,y)=\frac{2}{N}\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}C(u)C(v)F(u,v)\cos\left[\frac{(2x+1)u\pi}{2N}\right]\cos\left[\frac{(2y+1)v\pi}{2N}\right].$$

The transform basis patterns of the two dimensional DCT in the case of 8x8 is shown as an example in Fig. 13.

After performing the DCT of a video signal, a significant portion of energy tends to be concentrated in the DCT coefficients in the low frequency bands, even if there is no statistical deviation in a pixel block itself. Therefore, coding is performed according to the human visual system and the statistical deviation in the DCT coefficient domain of an image signal. An example of an image after transformation by DCT is shown in Fig. 14.

DCT coefficients are encoded by using zigzag scan and run length coding technique after quantization. Run length coding is a method of coding the combination of (number, length) of the same kinds of continuous symbols. Higher power DCT coefficients tend to be concentrated in the low frequency bands and the power becomes lower, even down to zero, as the frequency increases. The quantized indexes obtained by quantization of the DCT coefficients are scanned in a zigzag pattern from the low frequencies (upper left) to the high frequencies (lower right) and are rearranged into a one dimensional series. The signal series is expressed as a pair of the number of zeros (zero run) and a non-zero value following the zero series (level). When the last non-zero value is reached, a special sign called EOB (End of block) is assigned to reduce coding signals. By following this process, the statistical nature of the signal series can be exploited. Namely, symbols that have a large level will typically have a short zero run and symbols that have a long zero run are typically associated with a small level. In this way, a variable length code can be assigned to the combination of (zero run, level) to be compressed with shorter codes assigned to more probable symbols and longer codes assigned to less probable ones. The example of a zigzag scan and run length coding adopted in MPEG-2 are shown in Fig. 15.

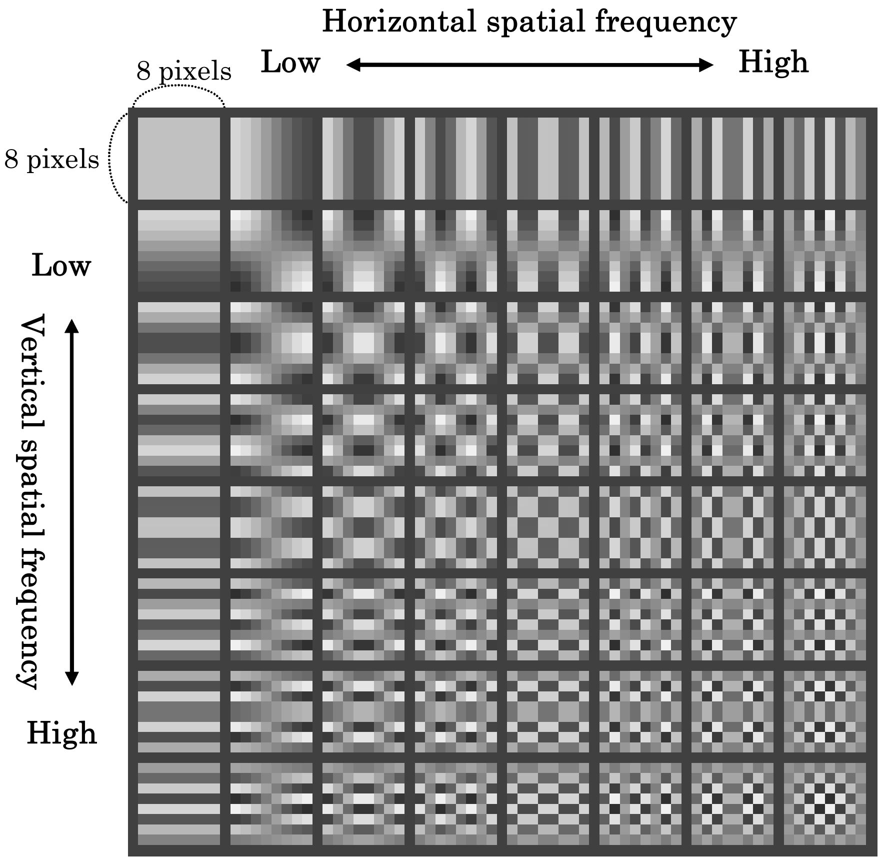


Fig. 13 Transform Basis Patterns of Two Dimensional 8x8 DCT

3.2.2.2 DWT

DWT is one of the transform methods using the transform basis made by the operation of expanding and moving a function localized in frequency domain. DWT allows using windows whose sizes are different according to frequencies, and has the feature of high response for both low-frequency and high-frequency portions of signals. DWT has also the following features,

(1) The correspondence for local waveform change is high by using flexible transform windows for unsteady signals
(2) Block noise which is often present in DCT transform does not occur inside the window width for the conversion of lowest frequency
(3) Hierarchical coding can be realized easily.

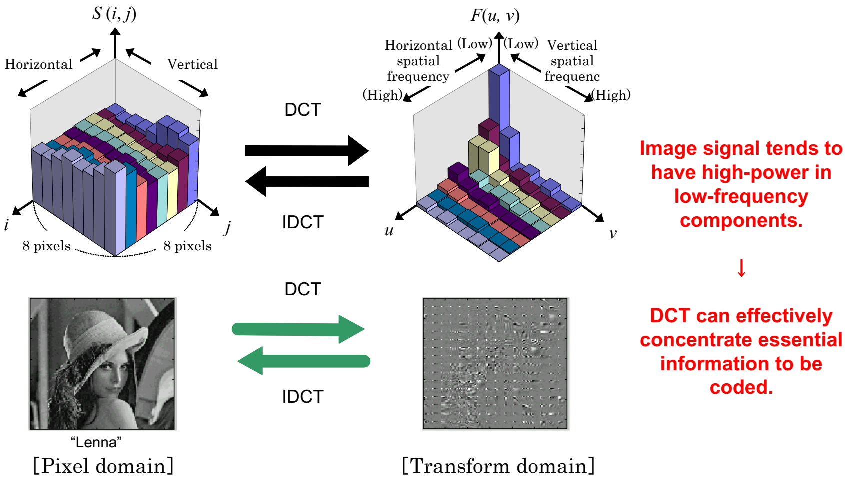


Fig. 14 Example of transformation by DCT

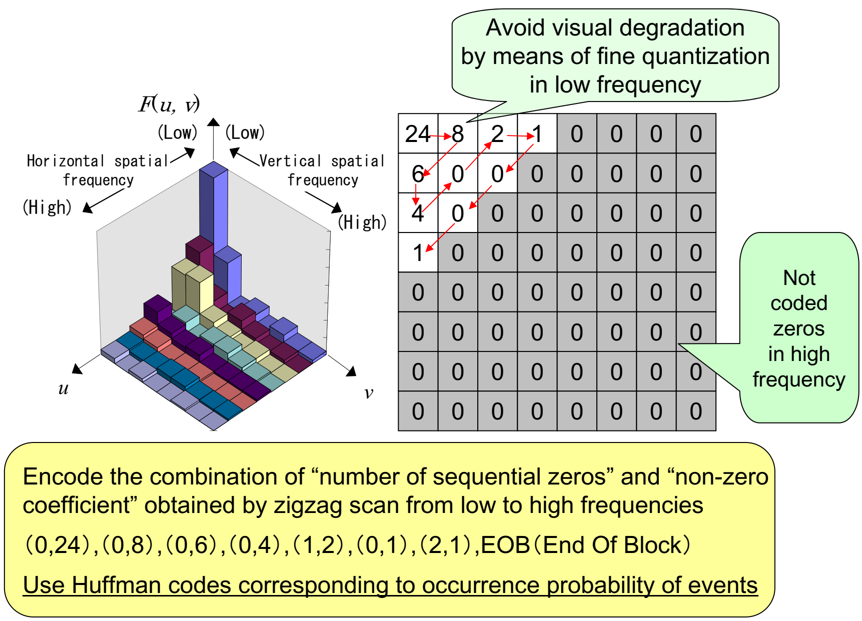


Fig. 15 Example of Zigzag Scan and Run Length Coding

3.2.3 Quantization

Quantization is a technique of reducing the amount of information directly, and there are mainly two methods well-known for video compression, which are Scalar Quantization and Vector Quantization. Scalar Quantization is an operation of making an input signal correspond to one of k kinds of values which are represented as q_1, ..., q_k as shown in Fig. 17.

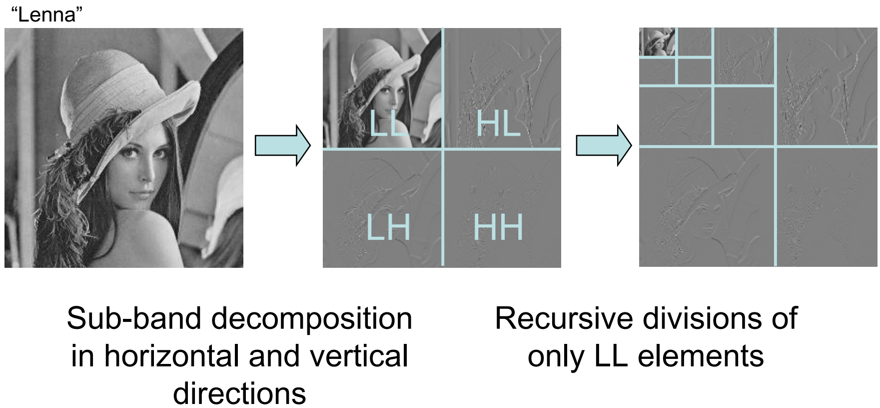


Fig. 16 Discrete Wavelet Transform

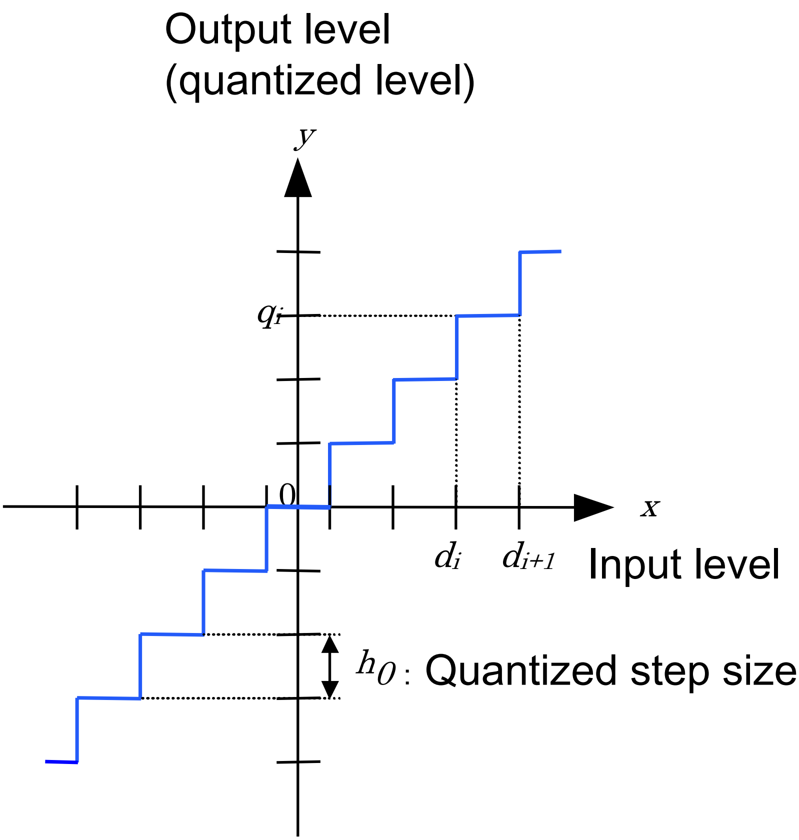


Fig. 17 Example of Scalar Quantization

Vector Quantization is an operation which quantizes several samples at the same time and expresses them with a representative vector which gives the best approximation of the samples [16, 17]. The sources of information which consist of many dimensions are quantized by one of the representative points of a multi-dimension space by Vector Quantization. Therefore, Vector Quantization has the following advantages,

(1) Coding efficiency can be raised by adopting the correlation and the dependency between the vectorized samples in the quantization mechanism.
(2) Even if the vectorized samples are completely independent, the multi-dimensional signal space can be divided into its quantized sections.
(3) Samples can be coded with non-integer word size by assigning a quantized represented vector or its codeword.

The key map and the principle of Vector Quantization are shown in Fig. 18 and Fig. 19, respectively.

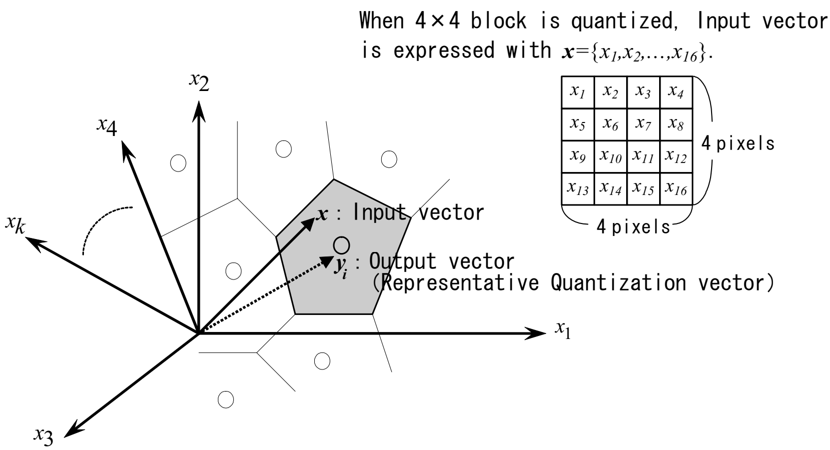


Fig. 18 Key map of Vector Quantization

3.2.4 Entropy Coding

Entropy Coding is a method of describing the mode information, motion vector information, quantized values, etc. as a series of binary signals which consists of only 0 and 1 (binarization). The total amount of codes is reducible by assigning coded words according to the occurrence probability of symbols. Huffman coding and arithmetic coding are typical entropy coding methods used in video coding. Huffman coding is a method of designing and using a variable length code table which associates symbols and code-words. This method can shorten the average code length by assigning short codes to symbols with high occurrence probability and long codes to symbols with low occurrence probability. An example of Huffman coding is shown in Fig. 20.

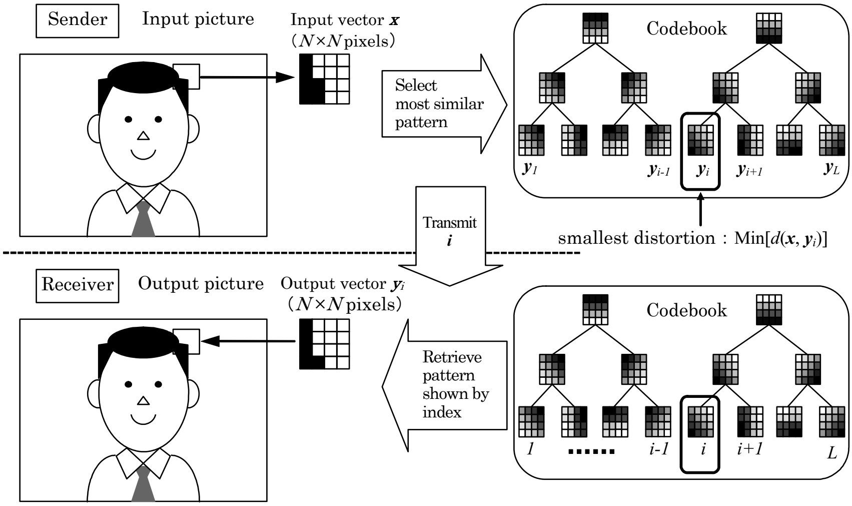


Fig. 19 Principle of Vector Quantization

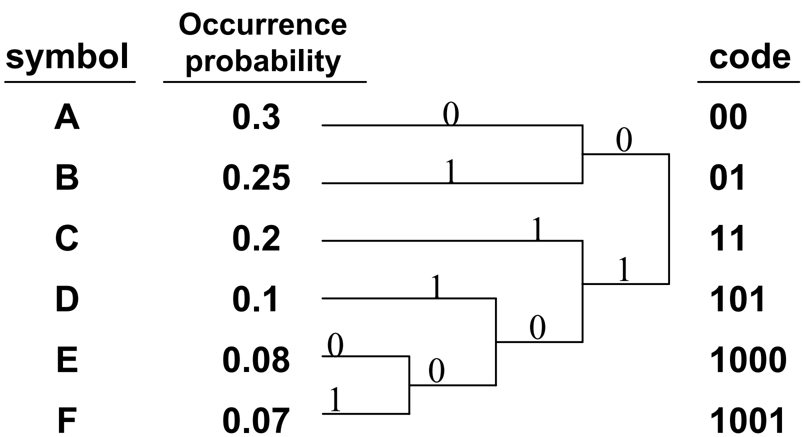


Fig. 20 Example of generation of Huffman Coding

Arithmetic coding is a method of coding the divided section of an interval of number line and generates a codeword one-by-one according to the occurrence probability of symbols. Moreover, the code length of a non-integer bit can be assigned to a symbol. The concept of Arithmetic coding is shown in Fig. 21.

3.2.5 Hybrid Coding Architecture

The Hybrid Coding Architecture which combines Prediction and Transform techniques is adopted in video coding standards such as H.26x and MPEG. The brock diagram of the typical Hybrid Coding Architecture is shown in Fig. 22.

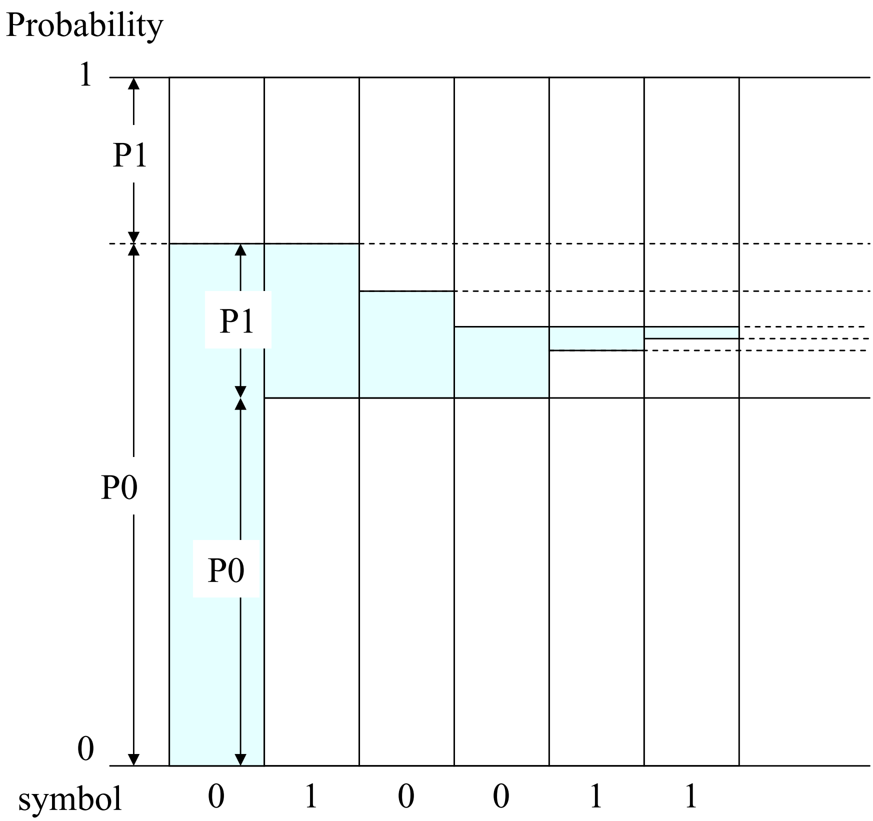


Fig. 21 Process of Arithmetic Coding

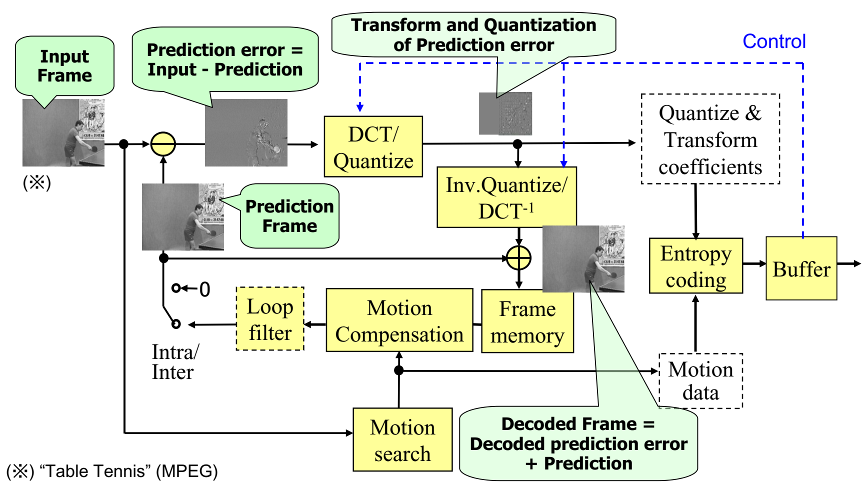


Fig. 22 Brock Diagram of Typical Hybrid Coding Architecture

3.3 MPEG Coding Methods

The features of the major MPEG video coding standards such as MPEG-2, MPEG-4 visual and AVC/H.264 are introduced [18, 19].

3.3.1 MPEG-2

MPEG-2 is an international standard that specifies video coding for the purpose of digital television broadcasting with high quality. Various coding tools are included to support the coding of interlaced video signals. MPEG-2 is used in satellite and terrestrial digital broadcastings and recording medias such as DVD, and is the mainstream coding method for video at present. The main features of MPEG-2 video coding are described below.

First, it adopts a hierarchical structure of video format. That is, MPEG-2 video format has a layered structure composed of Sequence, Group of Picture (GOP), Picture, Slice and Macroblock (MB) as shown in Fig. 23.

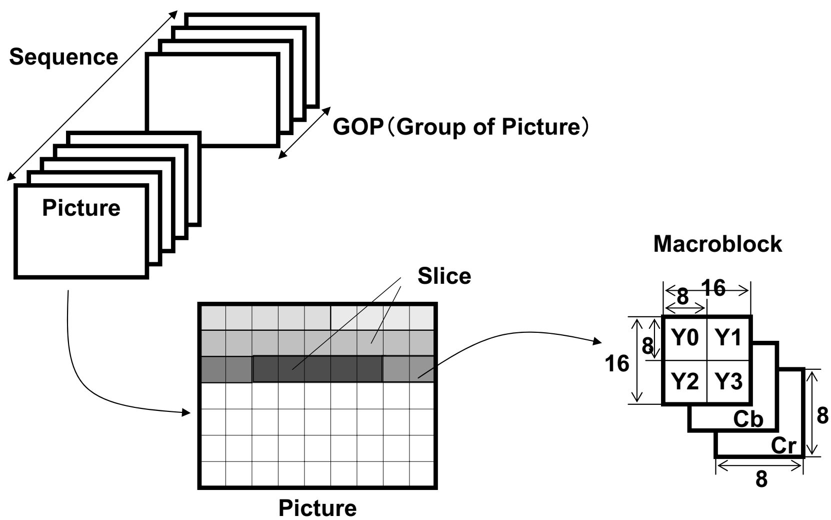


Fig. 23 Hierarchical structure of MPEG-2 video format

Secondly, three kinds of fundamental picture types and various prediction methods are adopted for Motion Compensated Prediction. Prediction efficiency is increased by adopting forward, backward, and bi-directional predictions. Forward prediction is a method of predicting the present frame from the past frame in time. Backward prediction is a method of coding the future frame and predicting a past frame from it. Bi-directional prediction is a method of using both the past and the future frames for prediction. Then, three kinds of pictures such as I, P, and B pictures are defined for using these prediction methods. I pictures are predicted within itself and do not refer to any other pictures. P pictures are predicted with only forward prediction. B pictures are predicted by choosing the most effective prediction among forward, backward and bi-directional predictions. A classification of picture types and the prediction methods of MPEG-2 are shown in Fig. 24. This structure makes it possible to realize random access of picture and also improves the coding performance for package media.

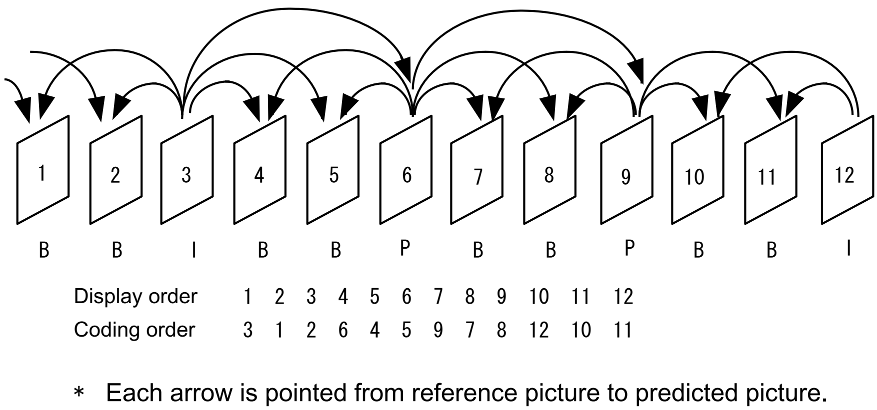


Fig. 24 Picture Classification and Prediction Methods

In addition, prediction with half-pel accuracy is defined in MPEG-2, whereby the unit of displacement is expressed with a motion vector pointing to half a pixel position (middle position of adjacent pixels). Since the value of half a pixel position does not actually exist, it is virtually generated by interpolation from neighboring pixels. The concept of half-pel accuracy prediction and the half-pel calculation method are shown in Fig. 25 and Fig. 26, respectively.

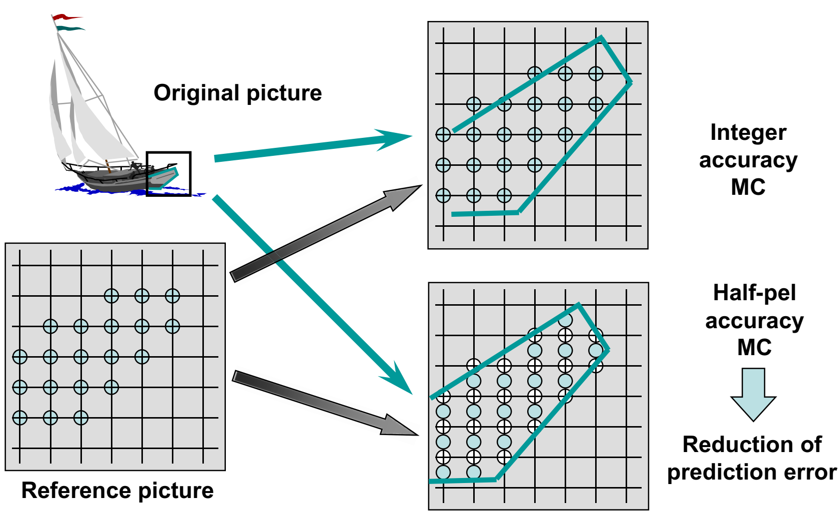


Fig. 25 Half-pel Accuracy Prediction

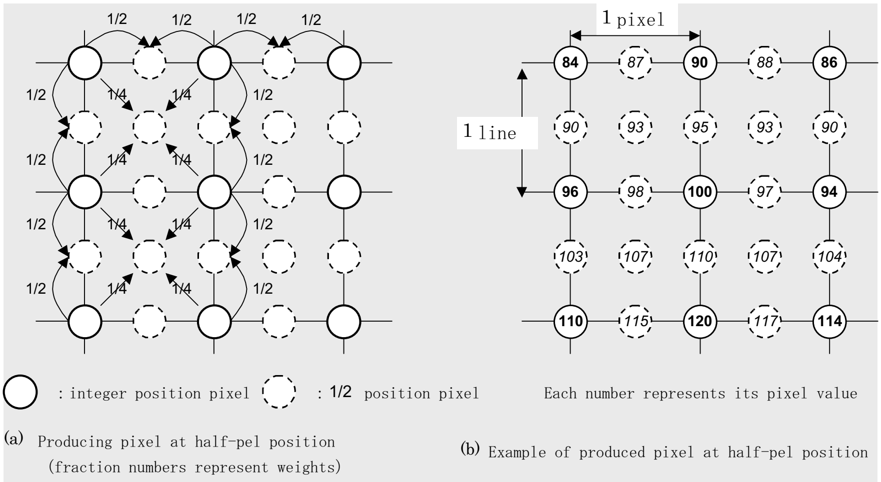


Fig. 26 Half a pixel Calculation Method

MPEG-2 has adopted various coding tools to support the efficient coding of Interlace video signals. First, Frame prediction, Field prediction or Dual prime prediction can be selected adaptively in order to perform optimal prediction according to the movement of the objects in video as shown in Fig. 27.

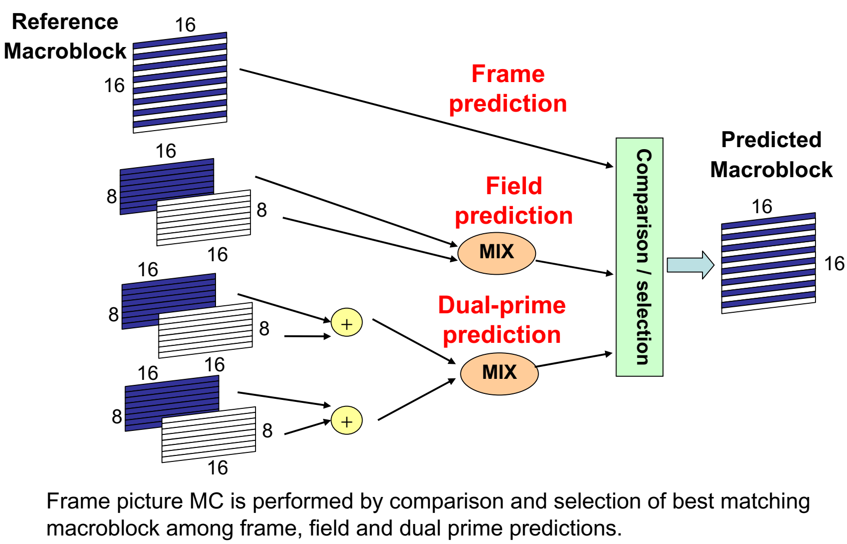


Fig. 27 Frame/field Adaptive Prediction

Moreover, Frame DCT or Field DCT can be chosen adaptively according to the video format as shown in Fig. 28.

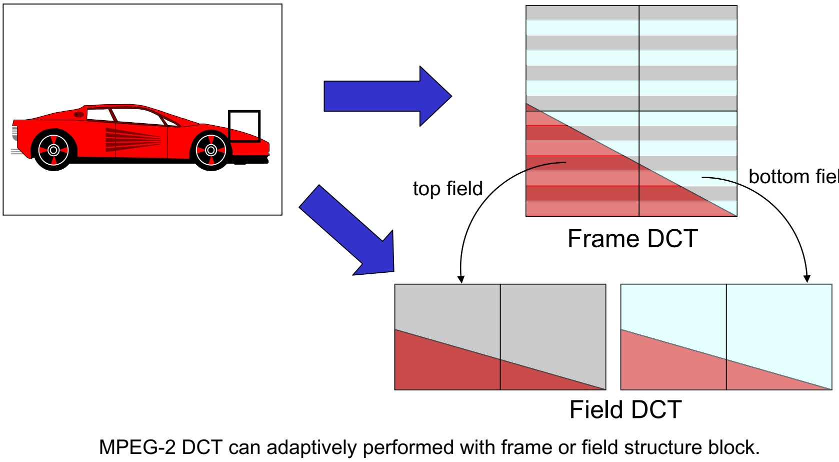


Fig. 28 Frame/field Adaptive DCT

Furthermore, scanning order can be switched adaptively according to frame/field DCT transform as shown in Fig. 29.

0	1	5	6	14	15	27	28
2	4	7	13	16	26	29	42
3	8	12	17	25	30	41	43
9	11	18	24	31	40	44	53
10	19	23	32	39	45	52	54
20	22	33	38	46	51	55	60
21	34	37	47	50	56	59	61
35	36	48	49	57	58	62	63

(a) zigzag scan

0	4	6	20	22	36	38	52
1	5	7	21	23	37	39	53
2	8	19	24	34	40	50	54
3	9	18	25	35	41	51	55
10	17	26	30	42	46	56	60
11	16	27	31	43	47	57	61
12	15	28	32	44	48	58	62
13	14	29	33	45	49	59	63

(b) alternate scan

Fig. 29 Adaptive scan change

The composition of MPEG-2 video coding is shown in Fig. 30.

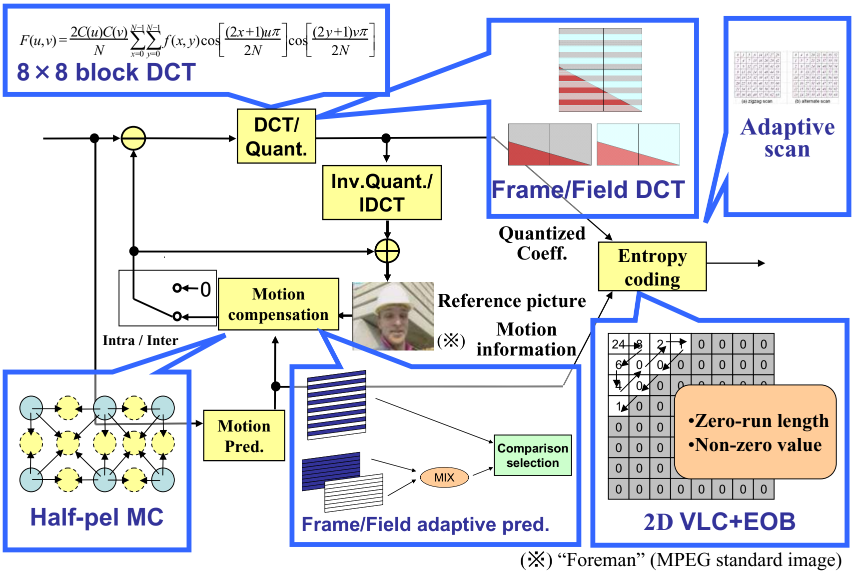


Fig. 30 MPEG-2 Video Coding

3.3.2 MPEG-4 visual

MPEG-4 visual is an international standard for the purpose of the coding at low bit rate for mobile equipments. It is used in mobile devices such as cellular phones and portable video players. MPEG-4 mainly performs coding and transmission for progressive video signal from QCIF (176 pixels x 144 lines) to VGA (640 pixels x 480 lines) at lower bit rates of about 1 - 3 Mbit/s. The main features of MPEG-4 visual are shown below.

First, Intra Frame Prediction is performed for DC and AC data of transformed and quantized coefficients. The entropy of a symbol which should be coded as DCT coefficients can be reduced by prediction since the DC coefficient is equivalent to the average value in a block and AC coefficients including low frequency harmonics have high spatial correlations. The outline of Intra Frame Prediction in MPEG-4 is shown in Fig. 31.

Next, in addition to half-pel accuracy of Motion Compensated Prediction, MPEG-4 also supports quarter-pel accuracy prediction which uses the virtual samples between half-pel pixels as a candidate of the prediction. Additionally, a 16x16 pixel macroblock domain can be equally divided into four 8x8 sub-blocks and Motion Compensated Prediction can be adaptively performed in a 16x16 macroblock unit or the unit of an 8x8 sub-block. With this technique, the performance of prediction for complicated motions within a macroblock can be improved. Furthermore, three dimensional VLC is performed on the quantization indexes after transform and quantization. This method includes the information (LAST), which indicates that the coefficient to be coded is the last non-zero coefficient in a block, into the set of (zero run, non-zero value). Then, the set of (LAST, zero run, non-zero value) is coded.

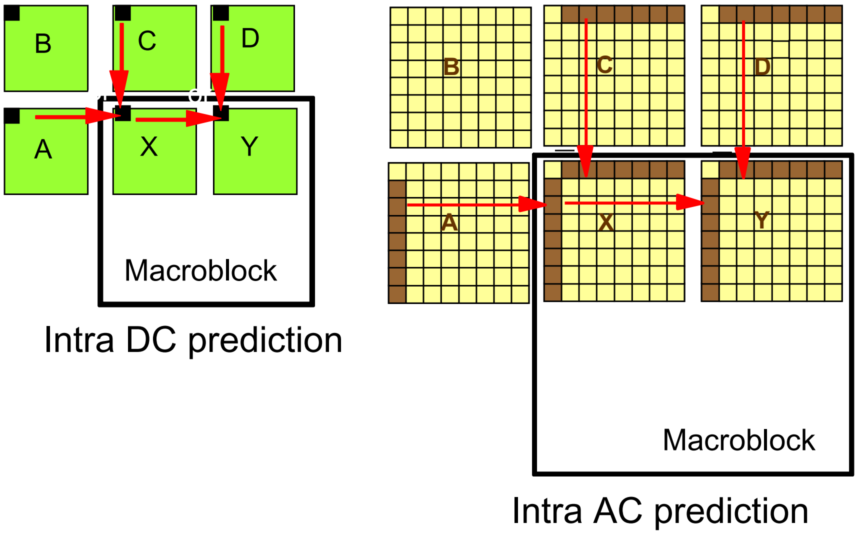


Fig. 31 Intra Frame DC/AC Prediction

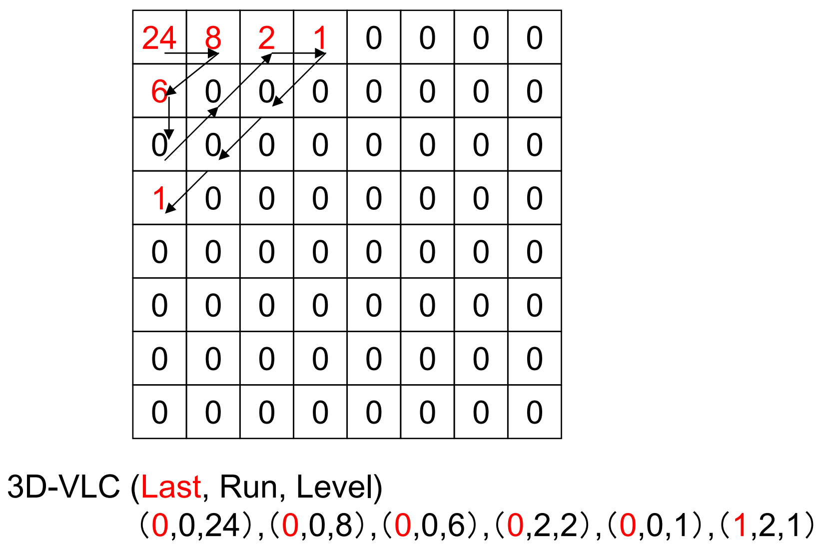


Fig. 32 Three dimensional VLC

The composition of MPEG-4 coding is shown in Fig. 33.

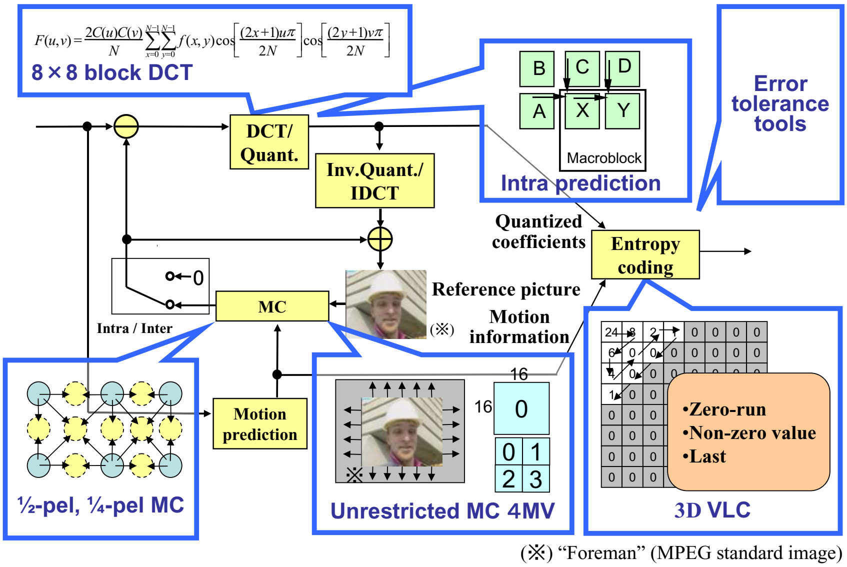


Fig. 33 MPEG-4 Visual coding

3.3.3 AVC/H.264

The AVC/H.264 standard is specified as MPEG-4 Part 10 by ISO/IEC as well as Recommendation H.264 by ITU-T. The improvement in coding efficiency is taken into consideration as the top priority when AVC/H.264 standardization was performed. It has been reported that AVC/H.264 has the twice as much compression efficiency of MPEG-2. In AVC/H.264, a multi-directional prediction in the spatial domain (pixel domain) is adopted as Intra Frame Prediction in order to reduce the amount of video information. Several prediction methods are defined for luminance and chrominance signals; 16x16 and 4x4 intra predictions for luminance are introduced below. Intra 16x16 prediction for luminance is a method which chooses either of four prediction modes shown in Fig. 34 per macroblock to predict a 16x16 pixel macroblock.

On the other hand, Intra 4x4 prediction for luminance divides a 16x16 pixel macroblock into 16 blocks which consist of 4x4-pixel blocks and chooses one of nine prediction modes as shown in Fig. 35 per block.

Moreover, an adaptive block size partition is adopted for Motion Compensated Prediction of AVC/H.264. Since seven block size partitions including 16x16, 16x8, 8x16, 8x8, 8x4, 4x8, and 4x4 are defined for Motion Compensation "Prediction, the size of prediction can be chosen per macroblock (16x16) or subblock (8x8). In addition, subblock partitioning can be used for four 8x8 blocks independently.

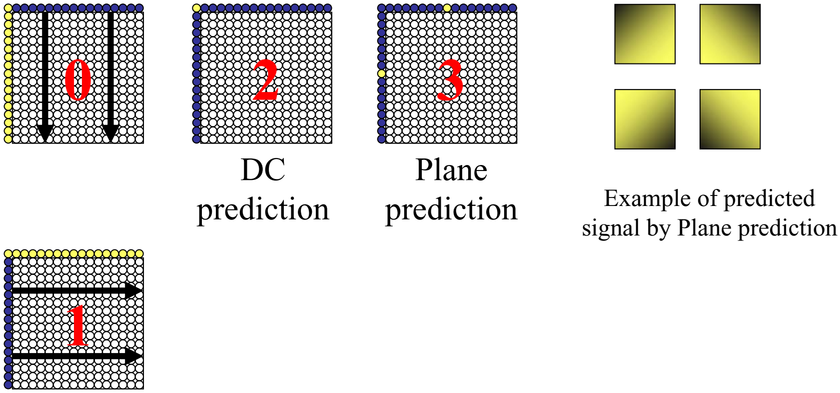


Fig. 34 Intra 16x16 prediction

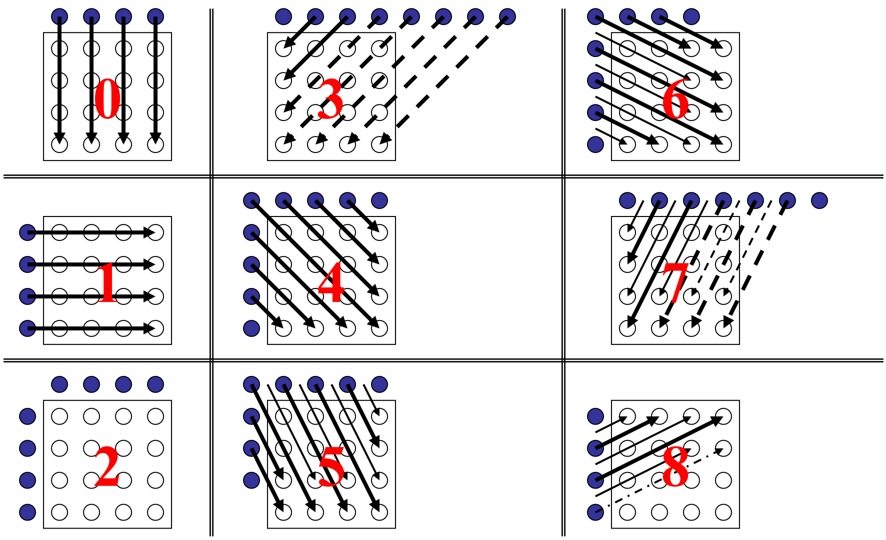


Fig. 35 Intra 4x4 prediction

In AVC/H.264, Motion Compensated Prediction can be performed by referring two or more reference frames. That is, the frames of the past and the future are stored in the frame memory, and can be chosen as reference frames for each block partition greater than 8x8 sub-blocks. B slices in AVC/H.264 support prediction from two reference pictures and the combination of the two pictures can be freely chosen. In contrast to MPEG-2, it is possible to perform bi-prediction even from two past pictures or two future pictures.

In AVC/H.264, 4x4 and 8x8 Integer Transform has been adopted for the conversion from spatial domain to frequency domain; the integer transform ensures that there is no mismatch between encoder and decoder. However, some parts of the transform process are included in quantization and de-quantization processing.

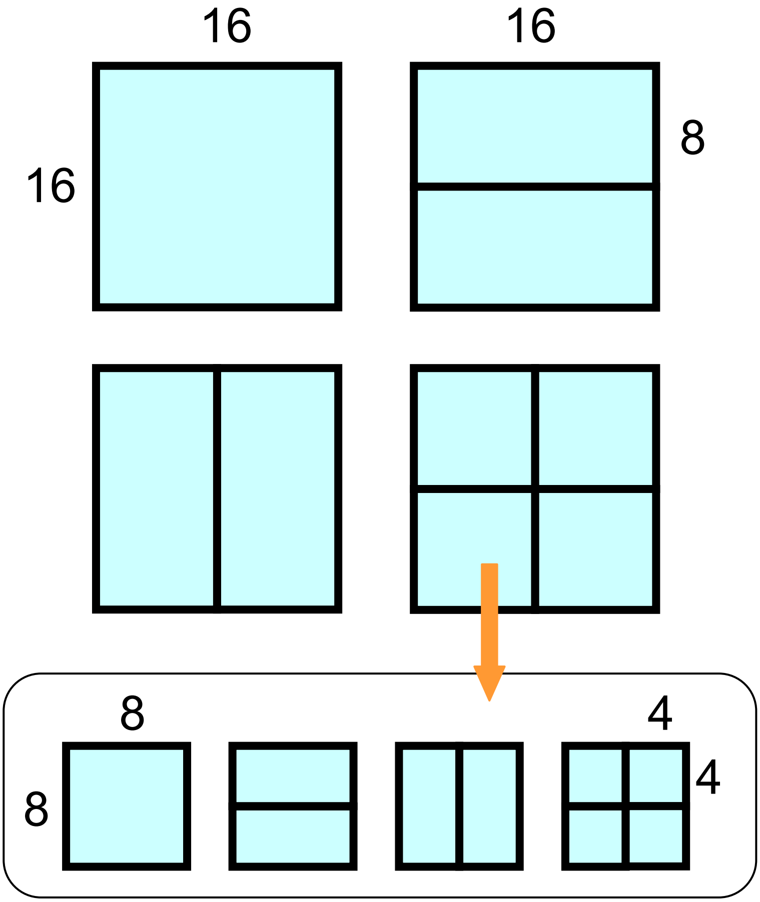


Fig. 36 Block size partitions of Motion Compensated Prediction

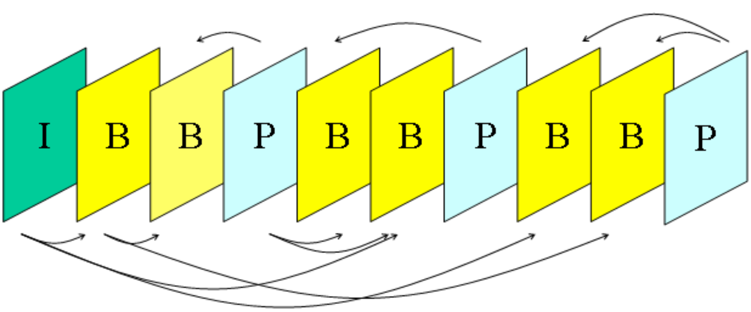


Fig. 37 Multi reference frame prediction

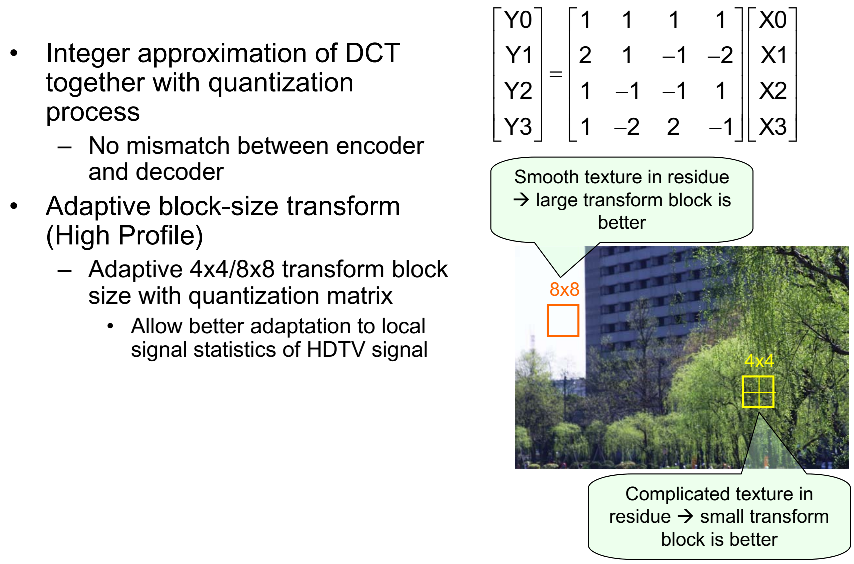


Fig. 38 Integer Transform

In entropy coding of AVC/H.264, high compression is achieved by encoding the symbols adaptively and using the knowledge (Context) in connection with coding states such as the information on the surrounding block data as well as variable length coding of coded symbols directly. Two types of entropy coding methods exist in AVC/H.264: Context Adaptive Variable length Coding (CAVLC) and Context Adaptive Binary Arithmetic Coding (CABAC). The CAVLC method is based on Huffman tables for encoding the symbols and generates contexts adaptively.

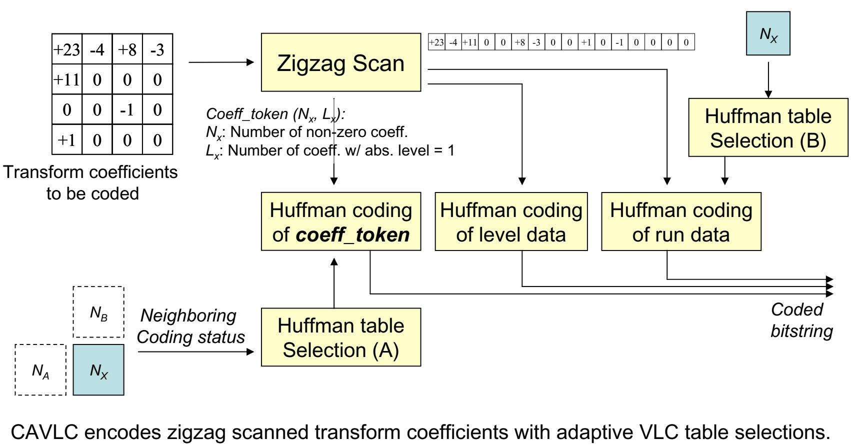


Fig. 39 CAVLC

On the other hand, CABAC converts symbols (coding mode, motion vector, transform coefficients, etc.) into a binary code series based on the rule defined by the standard (binarization), then chooses an occurrence stochastic model based on a context model (context modeling) and finally performs a binary arithmetic coding based on the selected occurrence stochastic model (binary arithmetic coding). In addition, an occurrence stochastic model is updated based on the result of coding (probability estimation).

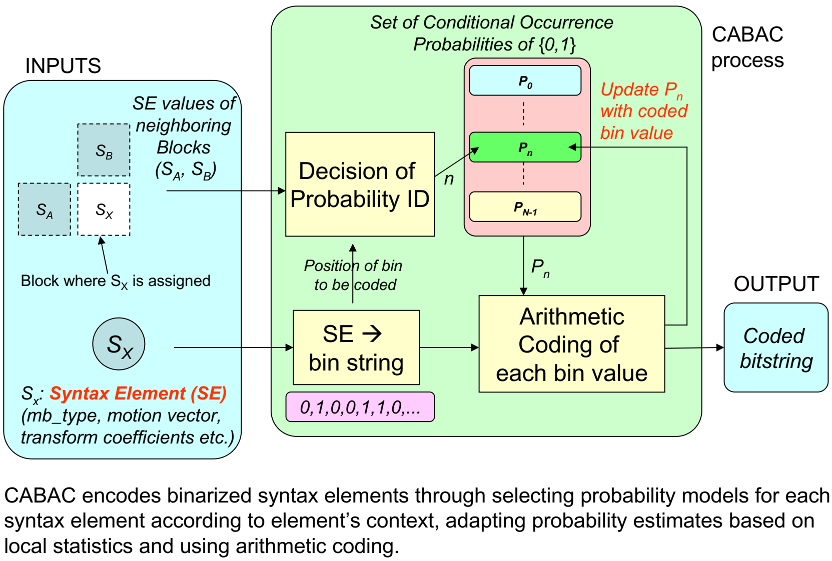


Fig. 40 CABAC

The composition of AVC/H.264 video coding is shown in Fig. 41.

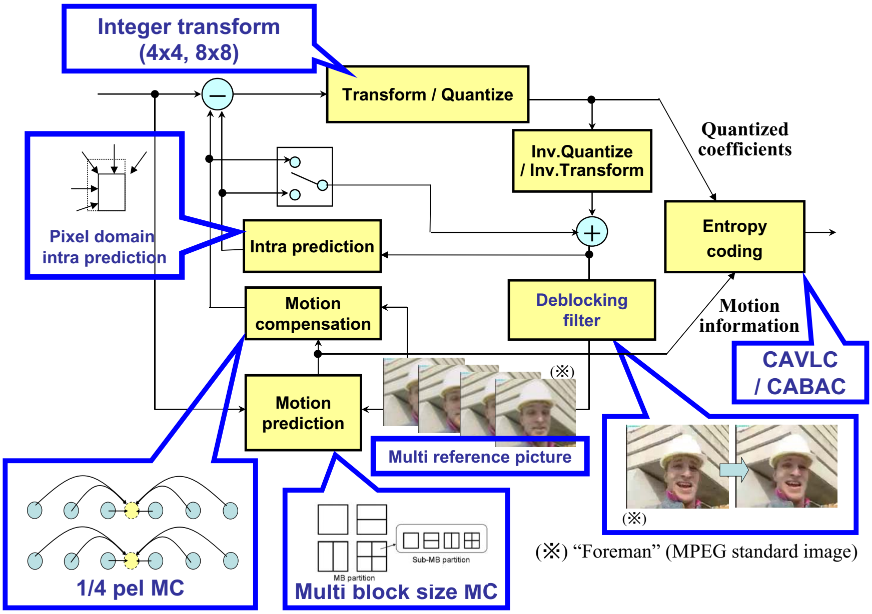


Fig. 41 AVC/H.264 video coding

3.3.4 Technical Achievements of MPEG Video Standards

The technical achievments of MPEG-2, MPEG-4 visual and AVC/H.264 are summarized in Table 1.

Table 1 Technical achievements of MPEG Video Standards

	Data Structure	Motion Compensation	Texture Coding (Intra and Inter)	Entropy Coding
MPEG-2	GOP including Interlace	I,P,B prediction / Half-pel MC / Interlace adaptation	Alternate Scanning / DCT mismatch control	2D Huffman
MPEG-4 visual	Video Object/ Sprite	16x16/8x8 MC / Quarter-pel MC / Unrestricted MC	Intra DC/AC Prediction Adaptive Scanning	3D Huffman
AVC/H.264	NAL Unit	16x16～4x4 Multi-blocksize MC / Multiple Reference Pictures	Pixel-domain Intra Prediction / Integer Transform / Adaptive Transform size	CAVLC / CABAC

Coding performance has been improved by the standard evolution of MPEG. The improvement of coding performance according to the progress of coding methods is shown in Fig. 42.

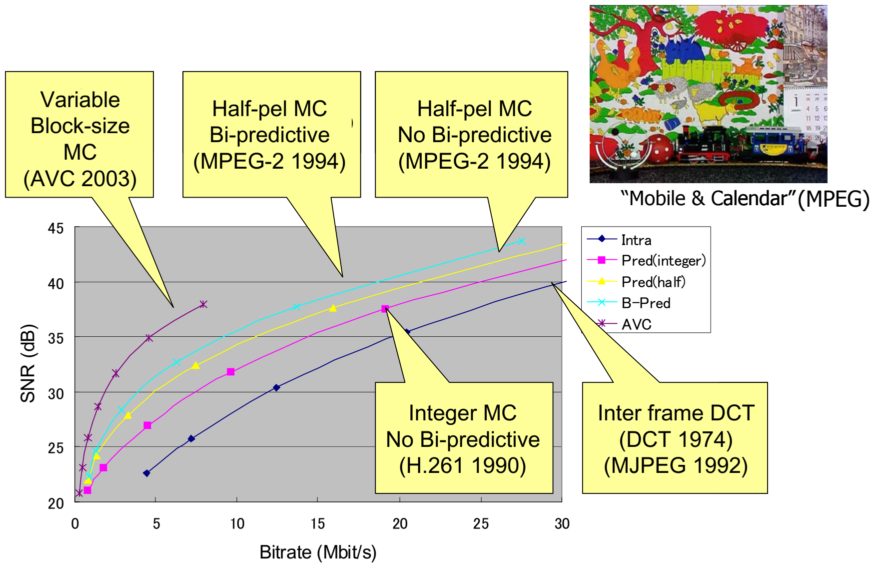


Fig. 42 Improvement of coding performance

On the other hand, the operation load in terms of complexity of video coding continues to increase with the standard evolution of MPEG. By improving Intra and Inter prediction accuracy, the complexity of AVC/H.264 encoding process has increased by 5 to 10 times compared with that of MPEG-2 or MPEG-4 visual. The complexity of the decoding process of AVC/H.264 is double for that of MPEG-4 visual by adopting context adaptive entropy coding and in-loop deblocking filter. However, the progress of semiconductor technology including LSI, processor and large scale storage has supported the realization of evolution of video compression technology.

4 Requirement for Quality of UHD Video System

4.1 Required Specifications for UHD Video Service

HDTV was realized by MPEG-2 as a digital broadcasting service for the home, and HDTV broadcasting has promoted both the thinness and enlargement of television displays. Camcorders and video recorders also support HDTV. Since the visible difference of video quality between SDTV and HDTV on a large screen became clear for users, the merits of HDTV have been validated. When UHD video, which is expected as the next generation video service, will be realized, it is necessary to improve each parameter, which influences video quality including spatial and temporal resolutions, gradation, and color space of HDTV. An illustration that re-examines the various factors to be qualified into UHD Video is shown in Fig. 43.

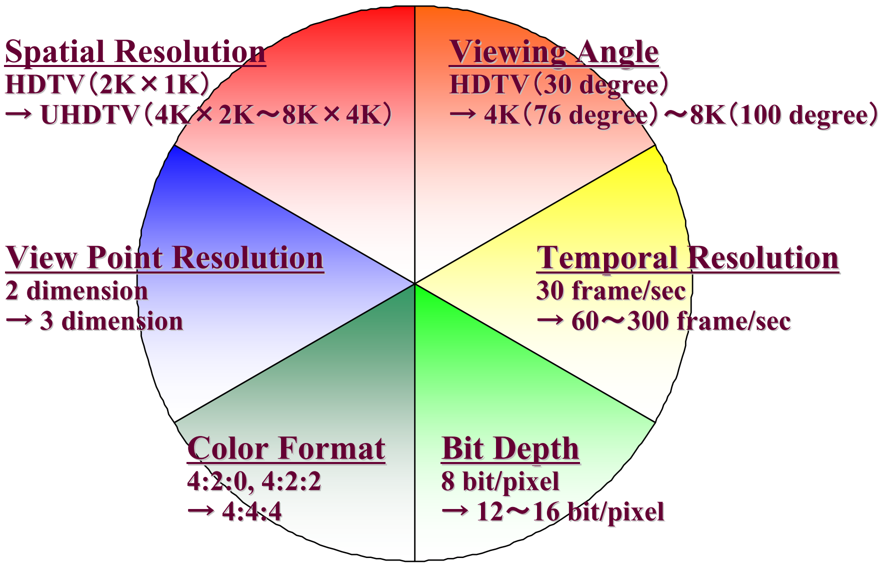


Fig. 43 Factors to be qualified into UHD Video

The specifications for UHD formats are planned to be standardized in ITU-R by 2012.

In the next section, the requirements about these resolutions and modes of the expression are considered, and the increase in the amount of information by fulfilling the requirements is measured.

4.1.1 Requirement for UHD Video

4.1.1.1 Spatial Resolution

Based on the fact that the angle of resolution of the human visual system is one minute degree, an HD image covers about 30 degrees of useful visual field, which can be recognized only with eyeball motion and without moving ones head. When the viewing angle is extended to 60 degrees of gazing viewing angle in which objects can be recognized only with little movement of ones head, 4K image can cover the angle. Furthermore, when a viewing angle is extended to 100 degrees of guidance viewing angle in which the existence of objects within the angle can be felt, 8K image can cover the angle. Although the video with 4K and 8K of UHD exceeding HD are already standardized as a video format of 3840x2160 and 7680x4320 in ITU-R Recommendation [20], video services with such resolution have not yet been deployed.

4.1.1.2 Temporal Resolution

30 fps (frame per second) used by the present television broadcasting was selected since it is the limit of the human visual system for flicker detection. However, it is

expected that at least 60 fps and up to 120 fps of temporal resolution is probably needed for UHD video since the present television broadcasting adopts an interlace signal and 60 fields per second is used in practice. If the temporal interval between frames is shortened, Motion Prediction of video coding will become more effective, and it should result in increased compression efficiency

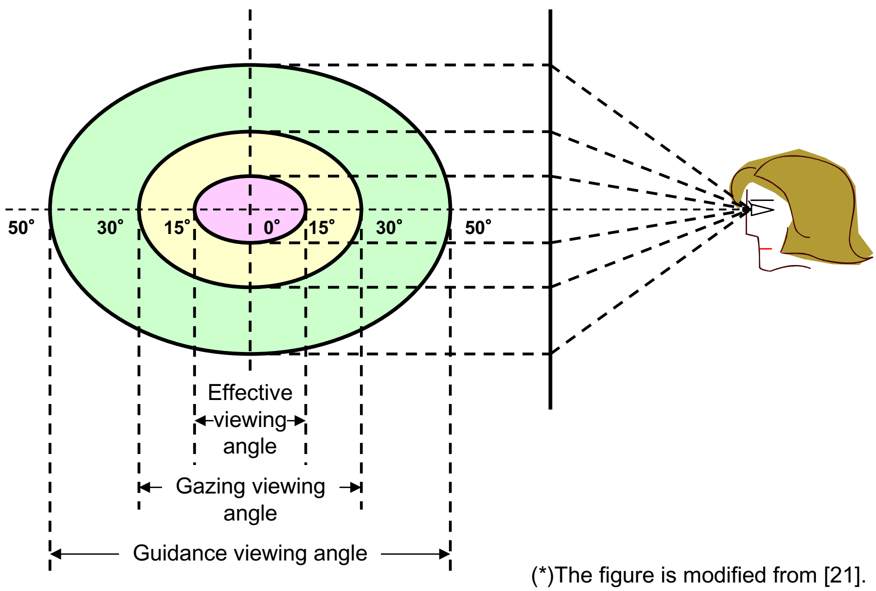


Fig. 44 The information acceptance characteristic within view [21]

4.1.1.3 Gradation per pixel

If the gradation per pixel increases, the contrast in the dark portions of a screen can be more visible. Moreover, there is also an advantage which can perform high precision calculations in the filter processing and the sub pixel processing for Motion Prediction in video coding. Therefore, the quality improvement of video is expected to be improved by increasing bit depth from the present 8 bpp to 10-12 bpp.

4.1.1.4 Color space

Although the amount of information becomes 1.5-2 times by increasing the color sampling format from 4:2:0/4:2:2 to 4:4:4, the possibility to use 4:4:4 format will increase if the compression efficiency of video coding is improved. Since the number of the signal elements of 4:2:0 format is one fourth of that of 4:4:4 format for chrominance signal, there is a trade-off in quality for video coding. At low bit rates, 4:2:0 video is preferred over 4:4:4 video since it contains fewer pixels. On the other hand, at higher bit rates, 4:4:4 video is preferred over 4:2:0 video. As a result, there is a cross-over point on performance between 4:2:0 and 4:4:4 video coding. For example, even if the cross-over point between 4:2:0 and 4:4:4 formats occurs at a higher bit rate than that of the practical use with existing video coding technology, the realization of high quality video applications with 4:4:4 format will be attained since the cross-over point is achieved to shift to a lower rate by the development of a new video coding technology with higher coding performance[22].

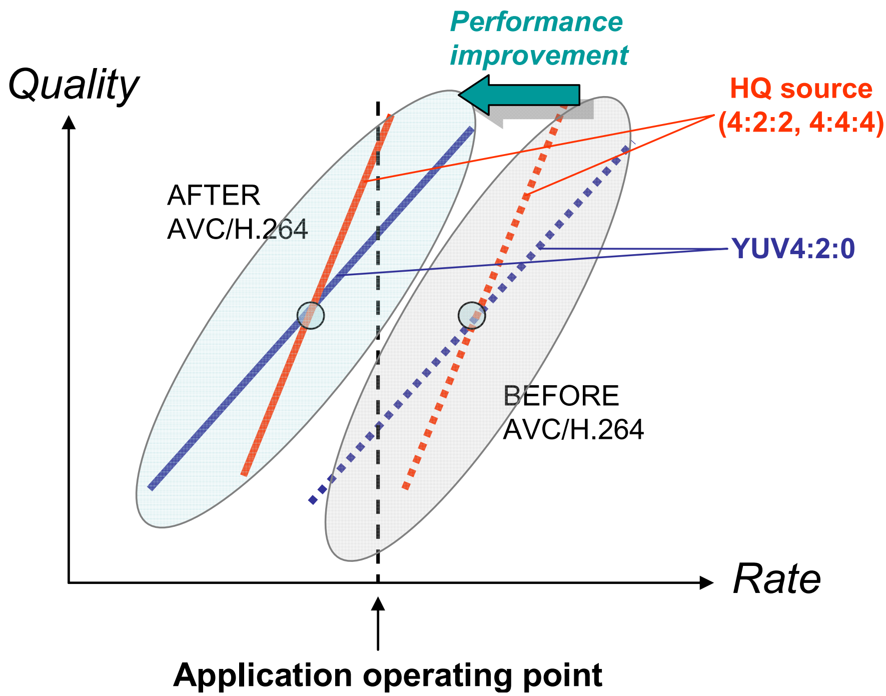


Fig. 45 Recent Progress of Video Coding Technology on Rate-Distortion Characteristic [13]

Moreover, most of video cameras and displays can treat RGB signal directly now. Then, if RGB and 4:4:4 formats are treated directly also in video coding, the degradation of the quality by means of color conversion does not occur and high quality video can be consistently provided.

4.1.2 The Amount of Information of UHD Video

Realizing new video expression which fulfills the requirements of UHD video is simultaneously accompanied with the steep increase of the amount of information. For example, the uncompressed rate of HD video (1920x1080/8bpp/4:2:0/30fps) is around 1 Gbit/s. This uncompressed rate increases to 3 to 18 Gbit/s for 4K video (3840x2160/8-12bpp/YUV4:2:0-4:4:4/30-60fps) and 12 to 72 Gbit/s for 8K video (7680x4320/8-12bpp/YUV4:2:0-4:4:4/30 - 60fps). Even if compared by the same bit length (8bpp), video format (4:2:2) and frame rate (30fps), 4K video has around four times more information and 8K has about 16 times more information than HD. With exponential increase of video data band width, the transmission rate for the interface between display/camera and storage/codec should also be exponentially increased as shown in Fig. 47. To satisfy the needs of I/O interface for 4K UHD Video, the standardization of 25Gbps serial optical interface for studio use is currently under consideration in SMPTE 32-NF-30.

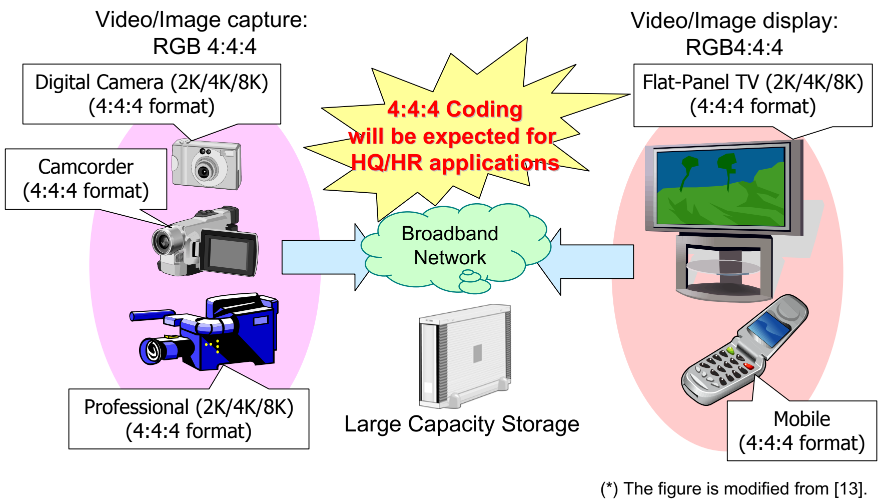
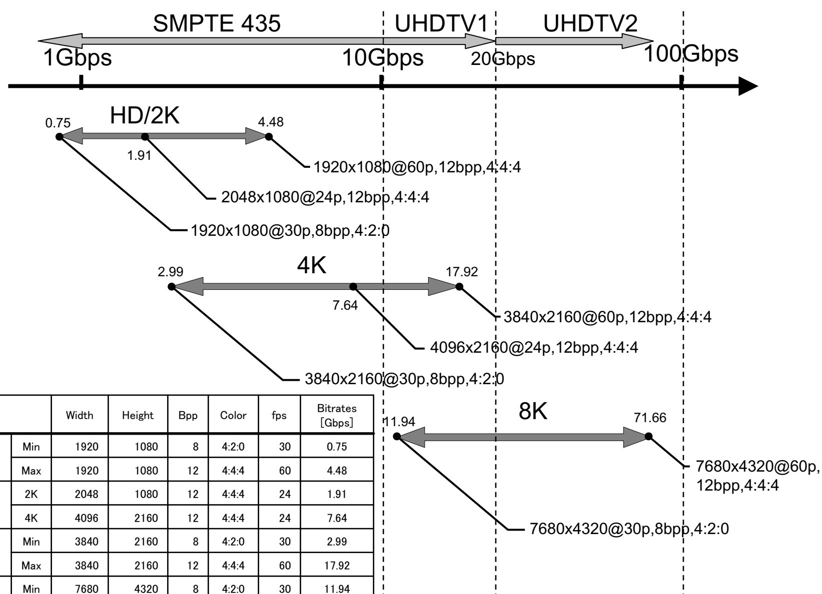


Fig. 46 High Quality and High Resolution/real-Color Video Applications [13]



		Width	Height	Bpp	Color	fps	Bitrates [Gbps]
HDTV	Min	1920	1080	8	4:2:0	30	0.75
	Max	1920	1080	12	4:4:4	60	4.48
Cinema	2K	2048	1080	12	4:4:4	24	1.91
	4K	4096	2160	12	4:4:4	24	7.64
UHDTV1	Min	3840	2160	8	4:2:0	30	2.99
	Max	3840	2160	12	4:4:4	60	17.92
UHDTV2	Min	7680	4320	8	4:2:0	30	11.94
	Max	7680	4320	12	4:4:4	60	71.66

Fig. 47 Bitrates for UHD Video Source [23]

4.2 Expectation for New Video Coding Technology

In order to utilize the UHD video with 4K and 8K resolutions, it is insufficient to use the present standard (AVC/H.264) which was standardized for the purpose of the coding video with the resolution up to HD. New video coding technology will be required, which compresses video greater while maintaining the high quality of an original video as much as possible. New video coding technology has a possibility to change SD/HD video compressed by MPEG-2 or AVC/H.264 into HD/UHD video and to exchange the television broadcasting which is the most familiar video media to the next generation TV.

5 Progress of Device Technologies Supporting UHD Video

There are several base technologies such as camera, display, storage, transmission system and video coding, which can realize next generation UHD video technology. There have been remarkable achievements in these areas until now. Therefore, we will soon be able to realize UHD video. In the following, the present status of camera, display, storage and digital network, is surveyed.

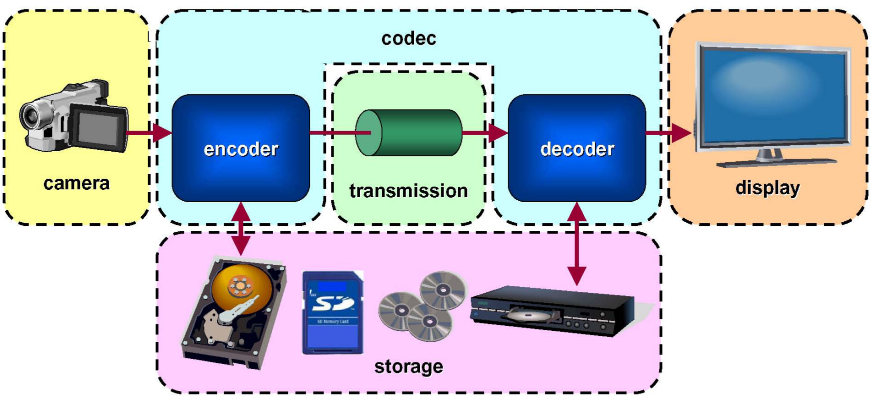


Fig. 48 Device technologies for realizing UHD video

5.1 Video Capture Device

Pixel size reduction of the image sensors for digital video cameras, in which high speed pick-up is possible, has progressed and the realization of high resolution cameras is ready.

With regards to consumer cameras, small camcorders that fit in the hand also support HDTV resolution with remarkably low price. Moreover, in professional use, several cameras with support of 4K resolution have already been announced and cameras with 8K resolution have also been developed. Since digital cinema services have started, the development of 4K cameras for digital cinema is progressing. For example, RED (RED Digital Cinema Camera Company) is manufacturing a 4K camera called RED ONE which is equipped with a CMOS sensor of 12 Mpixel and has the resolution of 4520x2540. DALSA Origin with 4096x2048 resolution (8 million pixels) has been developed by DALSA and the camera is characterized with 16 bit/pixel of high gradation. On the other hand, Octavision with 3840x2160 resolution (8 million pixels) has been developed by Olympus. A CMOS 4K camera which has a frame rate of 60fps with the resolution of 3840x2160 has been announced by Victor. Furthermore, an 8K video camera with 7680x4320 resolution (33 million pixels) captured by a high speed CMOS sensor, 60fps and 12 bit/pixel has been also announced by the Japan Broadcasting Corporation (NHK).

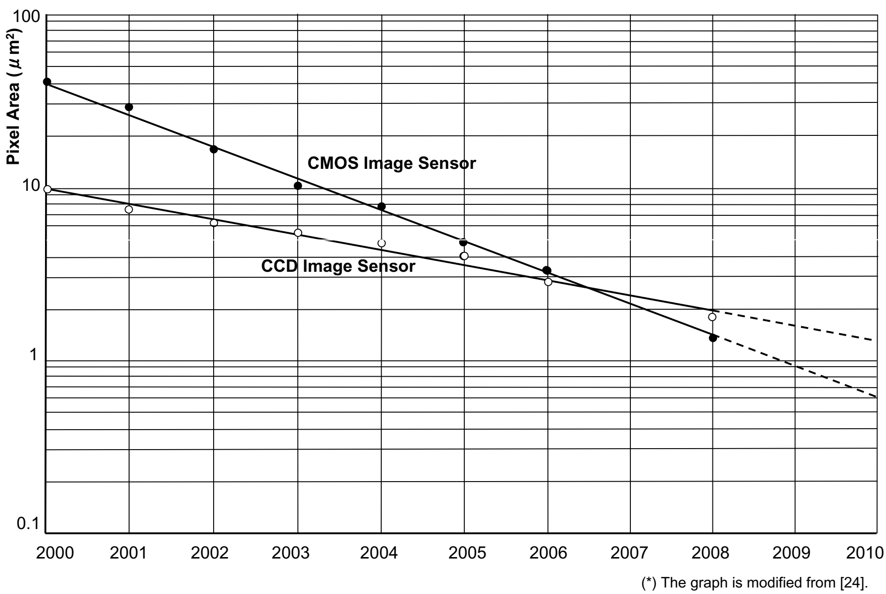


Fig. 49 Transition of image sensor pixel improvement [24]

5.2 Display

Flat panel TVs such as LCD and PDP have been progressing with larger screen and higher resolution. These TVs corresponded to the spread of digital contents including digital broadcasting. A screen size of the 40 inches has become popular and support for full HD resolution which can display images from television broadcasting is becoming typical.

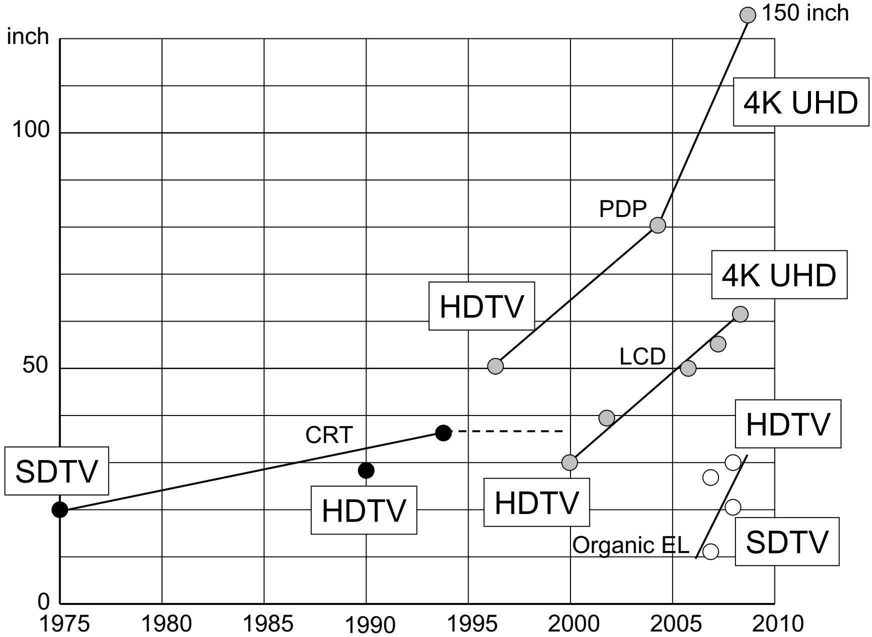


Fig. 50 Progress of TV screen size in Japan [25]

Moreover, the development of 4K television with 4 times as many pixels compared to full HD is progressing steadily and the possibility of its appearance in the home from 2010 to 2012 has also increased. There are trial products of LCD and PDP with 4K resolution including 3840x2160 and 4096x2160. The former expands Full HD twice horizontally and vertically, while the latter corresponds to digital cinema resolution. The LCD, PDP and projector which can display 4K video have been exhibited at several shows, and some of them are produced commercially. Several examples of 4K displays and 4K/8K projectors are shown in Table 2 and Table 3.

Table 2 Example of 4K and 8K digital cameras

Product, Manufacturer	Resolution	Frame Rate (fps)	notes
RED One	4520×2540	~30(4K) ~60(3K) ~120(2K)	2007~ CMOS, RAW data 12bit/pixel
DALSA Origin II	4096×2048	~30	2003~(Origin), 2007~(Origin Ⅱ) CCD, RAW data, 16bit/pixel
Olympus Octavision	3840×2160	24, 30	2005~ HDTV CCD, 4:2:2 format
JVC-Victor	3840×2160	60	2009 CMOS, RAW data 12bit/pixel
NHK	7680 x 4320	60	2009~ CMOS, RAW data 12bit/pixel
Vision Research Phantom 65	4096 x 2440	~125	2006~ CMOS

Table 3 Examples of 4K liquid crystal and plasma display

Manufacturer	Resolution	Thickness (inch)	System
SAMSUNG	3840 x 2160	82	Liquid Crystal (2008)
SAMSUNG SDI	4096 x 2160	63	Plasma (2008)
ASTRO design	3840 x 2160	56	Liquid Crystal(2007)
SHARP	4096 x 2160	64	Liquid Crystal (2008)
Panasonic	4096 x 2160	150	Plasma (2008)
NHK+Panasonic	3840 x 2160	103	Plasma (2009)
MITSUBISHI	3840 x 2160	56	Liquid Crystal(2007)

On the other hand, organic EL equipped with thinness, power saving and high resolution is very promising as a UHD video display. Organic EL displays have also been introduced at the various trade shows and some of them are produced commercially. Several organic EL displays are shown in Table 5.

Table 4 Examples of 4K and 8K Projectors

Product, Manufacturer	Resolution	notes
Victor DLA-SH4K	4096 x 2400	D-ILA(2007)
MERIDIAN	4096 x 2400	D-ILA(2009) House Use
SONY SRX	4096 x 2160	SXRD(2005)
NHK＋Victor	7680 x 4320	D-ILA(2004)
Victor	8192 x 4320	D-ILA(2009) RGB 12bit/pixel 60fps

Table 5 Examples of organic EL displays

Product, Manufacturer	Resolution	Thickness (inch)	notes
SAMSUNG	1920 x 1080	31	(2008)
SAMSUNG SDI	1920 x 1080	40	(2005/2008)
LG	1280 x 720	15	0.85mm (2009)
SONY	1920 x 1080	27	10mm(2007)
SONY XEL-1	960 x 540	11	3mm, (2007) ¥200,000

5.3 Storage

The speed at which storage capacity increases for HDD, SSD and other card type storage is remarkable. Storage capacity of a Terabyte has already arrived. Fig. 51 and Fig. 52 shows the improvement of recoding bit rate and strange density of typical storage media.

In the space of SD memory cards, SDHC with a capacity of 32 GB and a speed of 25MB/s has been realized, and SDXC with a capacity of 2 TB and a speed of 50-300MB/s has also been realized.

5.4 Digital Network for Communication and Broadcasting

The performance in the speed of optical line has progressed and some transmission experiments of the 4K resolution uncompressed video have been also

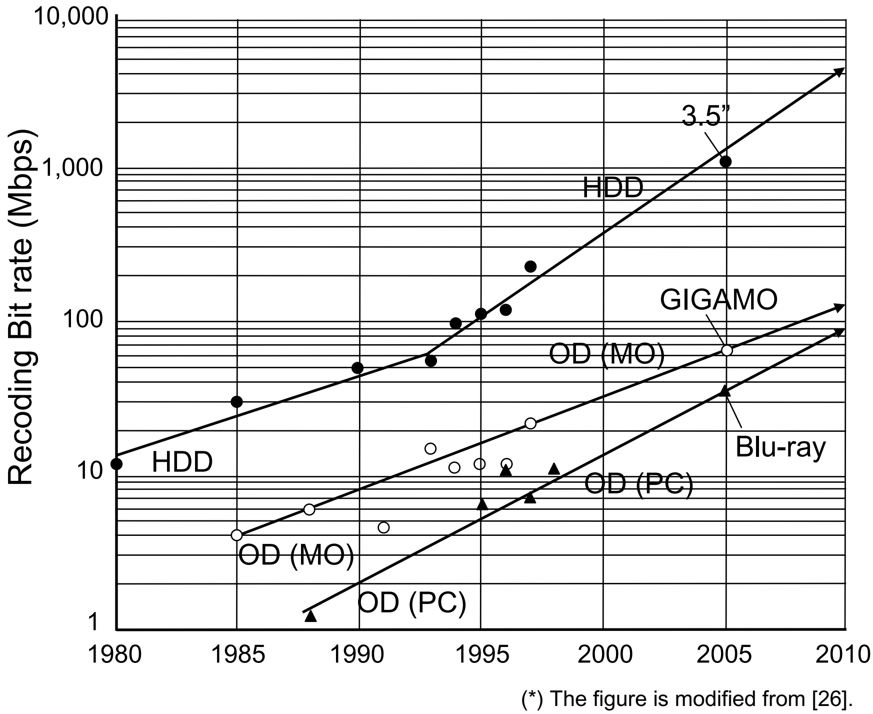


Fig. 51 Trend of Recording bit rate of hard disk and optical disk [26]

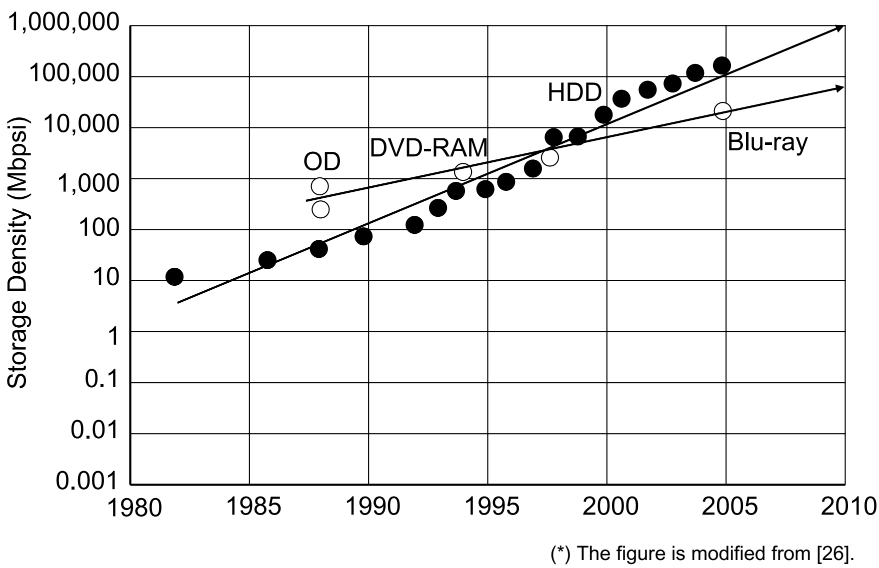


Fig. 52 Trend of Storage Density of Hard Disk and Optical Disk [26]

conducted. In Japan, the commercial service of broadband Ethernet with 1 Gbit/s has already been carried out, and a future service of 100 Gbit/s is also planned. Furthermore, UHD video will be able to be sent via broadband by NGN.

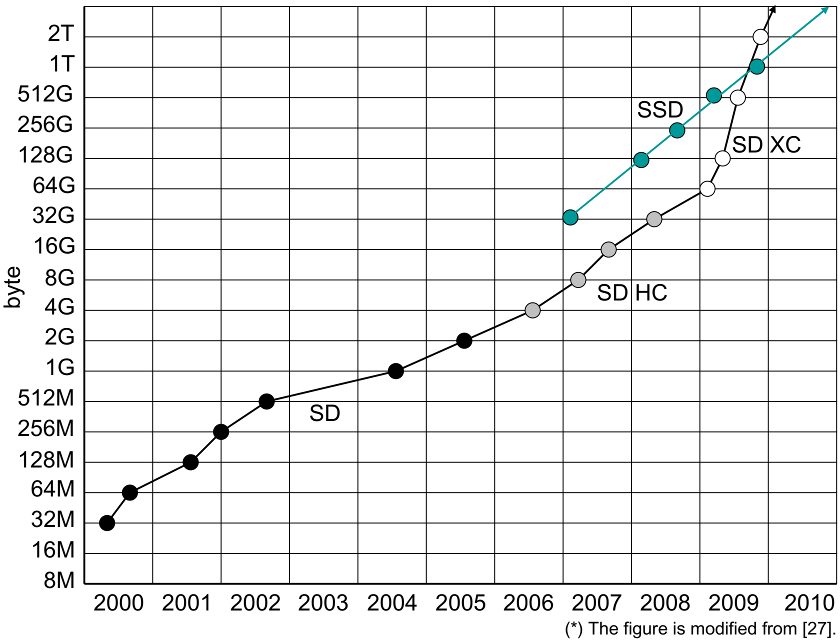


Fig. 53 Memory size of SD and SSD [27]

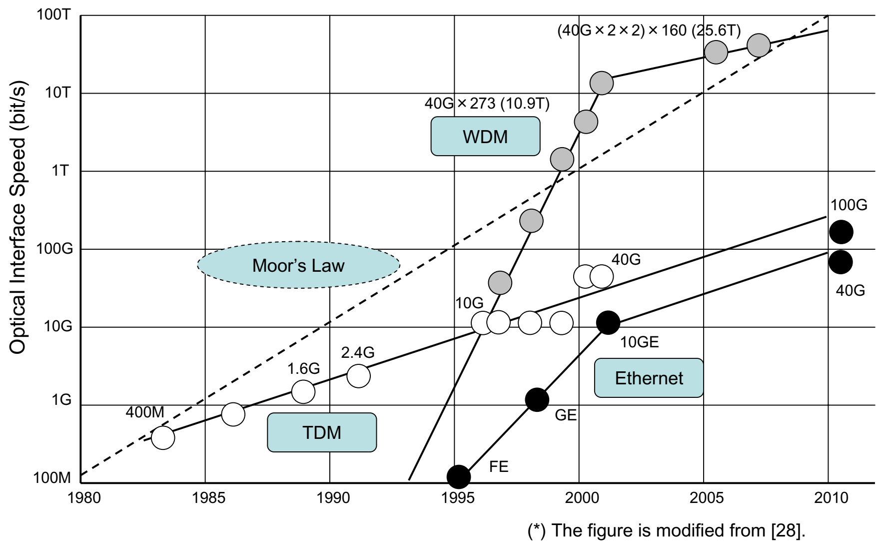


Fig. 54 Advance of optical transmission technology [28]

6 Standardization of UHD Video Coding Technology

6.1 Realization of UHD Video Application

When UHD video is considered, video coding technology is required to reduce the huge amount of information. An improvement in a compression ratio is needed while maintaining a high quality and ultra high resolution. Furthermore, from the scale and the cost of the video coding technology realized, UHD video technology is required to be a system which is more conscious of implementation. UHD video requires different inputs and outputs of a sensor and a device compared to present systems. Therefore, it is necessary to take the characteristics of those devices into consideration. Even if UHD video technology follows the framework of existing video encoding methods, its high resolution may change the optimal coding parameters. This change may result in the replacement of some coding tools in the current system. There is also the possibility that techniques which could be realized only in simulation are not able to be utilized in practical codecs because of memory integration, calculation complexity and device cost.

6.2 Next Generation Video Coding Standard

JVT have moved the principal axis of their activities toward the addition of other functionalities such as SVC (Scalable Video Coding) and MVC (Multi-view Video Coding) after the standardization of AVC/H.264 High Profile. However, since the development of UHD video devices has become remarkable from 2008 to 2009 and the necessity for UHD video coding was appealed from Japan, the argument of the standardization of a next generation video coding has become active toward the development of HVC (High performance Video Coding) in MPEG since April 2009. Then, MPEG invited the public to submit evidence of new video coding technologies that fulfill the conditions for UHD video, and evaluated the proposed technologies at the meeting in June-July 2009 [11]. As a result of this study, MPEG decided to work towards issuing a formal call for proposals and initiating the development of a new video coding standard. The current plan is to collect proposals in 2009, and to issue an international standard in 2012 or 2013. At the same time, VCEG is considering the improvement of video coding efficiency and the reduction of the complexity based on AVC/H.264 as KTA (Key Technical Areas). KTA has started to be discussed since 2005 and has continued to be added new coding tools. Then, the examination of NGVC (Next Generation Video Coding) was started in 2009, which is assumed to be used for UHD video, HD broadcasting, HD video conference, mobile entertainment, etc. A draft document describing the collection of test sequences and the requirement of NGVC was created during the Yokohama meeting in April 2009. The requirement was updated and the working title was changed from NGVC to EPVC (Enhanced Performance Video Coding) during the Geneva meeting in July 2009. In addition, the cooperation between MPEG and VCEG is likely to be established. The requirement conditions of MPEG HVC [29] and VCEG EPVC [30] are shown below.

(1) Video coding performance
To realize 50% reduction of coding bits with subjective quality equivalent to AVC/H.264 High Profile.
(2) Correspondence to a high quality video
(2.1) Resolution
From VGA to 4K (also 8K)
(2.2) Chroma format
From 4:2:0/4:2:2 to 4:4:4
(2.3) Bit length
8 - 14 bit per pixel
(2.4) Frame rate
24 - 60 fps or more
(2.5) Complexity
(i) To be possible of implementation at its standardization period.
(ii) To be possible to control the trade off between complexity and coding performance.
(iii) To be possible of parallel processing.

MPEG HVC and VCEG EPVC are going to publish a joint Call for Proposals on the next video coding standard in January 2010, and the target bitrates for UHD and HDTV are currently defined as shown in Table 6.

Table 6 Target bitrates for Call for Proposals on HVC/EPVC

Class	Resolution	Frame rate	Target bitrate for evaluation
A	2560 x 1600p cropped from 4K	30 [fps]	2.5[Mbps], 3.5[Mbps], 5.0[Mbps], 8.0[Mbps], 14.0[Mbps]
B	1920 x 1080p	24 [fps]	1.0[Mbps], 1.6[Mbps], 2.5[Mbps], 4.0[Mbps], 6.0[Mbps]
		50/60 [fps]	2.0[Mbps], 3.0[Mbps], 4.5[Mbps], 7.0[Mbps], 10.0[Mbps]

6.3 Challenge toward UHD Video Coding

6.3.1 Trial to improve AVC/H.264

Within VCEG, experiments have shown the improvement in video coding performance of AVC/H.264 using KTA and continual improvements of the reference software. Many proposals for Evidence collection of MPEG used the tools extracted from KTA. Therefore, it can be said that KTA at present is a benchmark for the improvement of video coding efficiency. The following tools have been incorporated into the KTA software and have been found to yield gains in coding efficiency.

(1) Extension of bit depth
(2) 1/8-pel accuracy Motion Compensated Prediction
(3) Prediction direction adaptive transform
(4) Adaptive prediction error coding
(5) Adaptive quantization matrix
(6) Adaptive quantization rounding
(7) RD optimal quantization
(8) Adaptive interpolation filter
(9) High precision interpolation filter
(10) Adaptive loop filter

The block diagram of encoder with these tools is shown in Fig. 55.

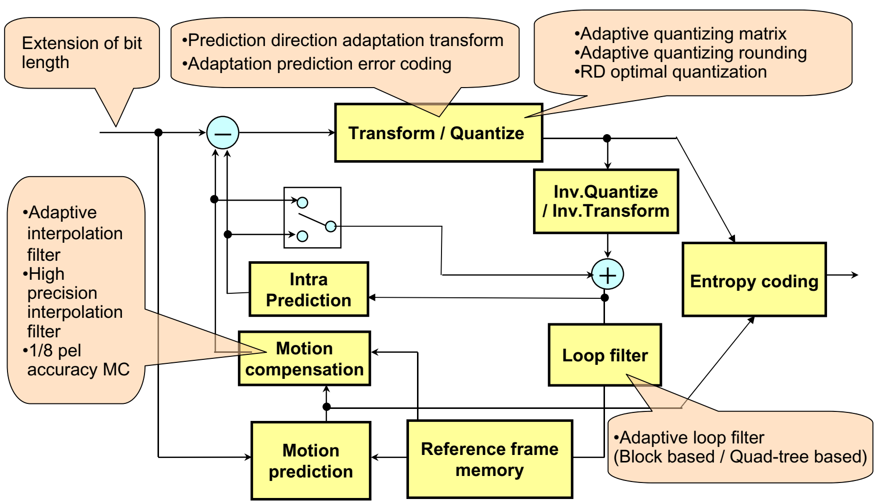


Fig. 55 Improvement points of AVC/H.264

6.3.1.1 Improvement of Coding Performance of KTA

Some new coding tools that have been incorporated into the KTA software that provide improvements in video coding efficiency for motion compensated prediction, transform, loop filter, and Intra prediction, are reviewed below.

(1) Motion Compensated Prediction

When generating the motion compensation signal of sub-pixel accuracy, a suitable filter can be chosen among several candidates and the interpolation method which increases operation accuracy is used. It is supposed that gains of about 4 to 7% of improvement are realized.

(2) Transform
The size of a macroblock is expanded to maximum of 64x64, and the transform of several block sizes from 4x4 to 16x16 is available. It is effective for high resolution video. It has been claimed that improvements in the range of 10 to 15% for P picture and 15 to 20% for B picture can be achieved.

(3) Loop Filter
The Wiener filter which performs image restoration from a local decoded picture using source picture is designed. The filter can be also turned on and off for each block to improve the quality of a local decoding picture. Gains of about 5 to 10% of improvement have been reported.

(4) Intra Coding
Prediction between pixels is alternatively performed from several directions in the Intra coding of AVC/H.264. In this tool, transform basis is changed according to the directions of the predictions. Improvements of about 7% have been shown.

These tools are the improvements to the existing AVC/H.264 framework and keep the existing coding architecture. They adopt the approach of changing the range of parameters, selecting adaptive case among several candidates and performing the optimization which could not be employed by the restriction of practical memory and operation scale before. It is said that KTA could provide about 20 to 30% of performance improvement relative to AVC/H.264 by adoption of these tools.

6.3.1.2 Block Size Extension and Non-Rectangle Block Application

A coding method which extends AVC/H.264 and performs UHD video coding using block size extension and a non-rectangle block is introduced as an example [31]. Macroblock size is extended from 16x16 to 32x32, and accordingly the block partitions for motion prediction are expanded to 32x32, 32x16, 16x32 and 16x16. When 16x16 block partition is chosen, sub-block partitions of 16x16, 16x8, 8x16 and 8x8 are also employed. Smaller block partitions of 8x4, 4x8 and 4x4 are not used because noise components will become dominant and the essential structure information of a picture will become impossible to be expressed efficiently by such small blocks in the case of UHD video coding.

In addition to extension of macroblock size, non rectangle block partitions are employed for motion prediction. Although non rectangle block partition was proposed for the purpose of expressing more complicated motion with a short motion vector in super low bit rate video coding, it is believed that new partitions will contribute to reducing residual energy by effectively expressing complicated picture structures such as object boundaries in UHD video. On the other hand, there are some problems such as the increase of the motion detection operations, the increase of the shape description information and the necessity of the memory access to complicated shape. Then, the simple diagonal partitions shown in Fig. 57 are adopted, which can be created with the combination of 16x16 blocks.

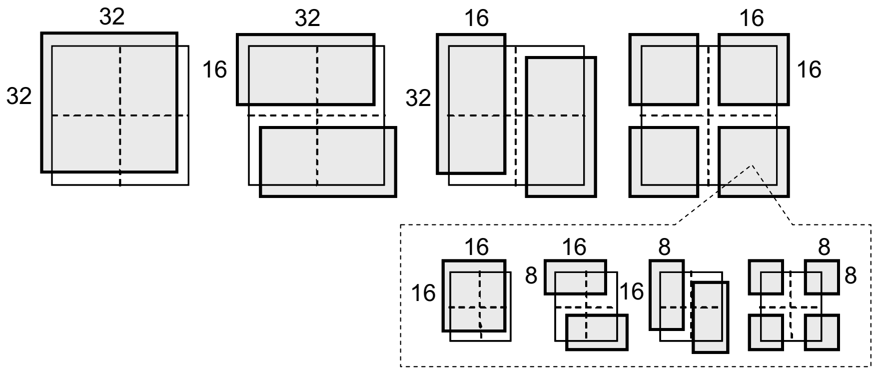


Fig. 56 Hierarchical division of motion compensated prediction blocks

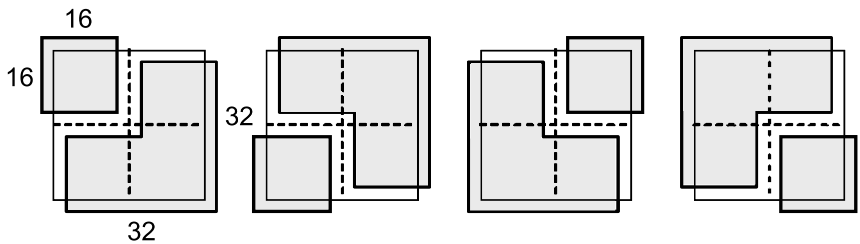


Fig. 57 Diagonal block partition

Moreover, correlation between pixels in wide range can be used if transform size is extended. Then, transform coefficients can be compacted more into the low frequency bands. Since it is preferred to not to include block boundaries in a block partition, it is effective to change the transform block size according to the block partition size for motion prediction. Then, 8x8 and 16x16 block sizes are adopted as transform block sizes and switched adaptively according to chosen motion prediction block partition. The above improvements were applied to P picture, and the computer simulation was performed using a GOP structure of IPPP. The conditions of the experiment are shown in Table 7. Reference software JM15.1 of AVC/H.264 was used as the anchor technique for the quality assessment. Moreover, the test sequences used for the experiment are shown in Fig. 58. Traffic of Class A and Kimono1 of Class B of MPEG test sequence (YUV 4:2:0 format and 8 bpp) were chosen for the experiments. Class A includes sequences with 2560x1600 resolution that have been cropped from 4096x2048 pictures; in this way the quality and compression efficiency of 4Kx2K could be practically evaluated. Class B includes the sequences of Full HD (1920x1080) size.

The results of the experiment show the improvement in efficiency of the motion compensated prediction and the decrease of addition information such as motion vector. The amount of coding bits is successfully reduced by 2 to 30% and PSNR was improved by 0.1-0.9 dB. The improvement of PSNR and the amount of

Table 7 Experimental condition

Test sequence (YUV4:2:0, 8bpp)	- Class A: (2560x1600) Traffic (30 fps) - Class B: (1920x1080) Kimono1 (24 fps)
Number of frames	- Class A: 300 frames - Class B: 240 frames
GOP composition	- IPPP - I picture interval : Class A: 28 frames Class B: 24 frames
MB size	- I picture: 16x16 - P picture Proposed: 32x32 Anchor: 16x16
Motion compensation prediction	- Motion search method: EPZS - Motion search range: ±128 pixels, 1/4-pixel accuracy - Intra/Inter switching ON - Block partition Proposed: rectangle and diagonal Anchor: only rectangle
Qp	- Class A: 25, 29, 33, 37 (fixed) - Class B: 25, 28, 31, 34 (fixed)
R-D optimization	ON
Entropy Coding	CABAC

coding bits are shown also in Fig 58. Moreover, the improvement of coding efficiency is depicted as a comparison of R-D curve with AVC/H.264 is shown in Fig. 59.

6.3.1.3 Other Possibilities of Performance Improvements for UHD Video Coding

In addition to the improvement of KTA tools and the expansion of the block size for UHD video, other possibilities of performance improvement of AVC/H.264 are assumed as follows.

(1) Examination of pre filter processing.
(2) The design of the effective probability table of arithmetic coding in CABAC.
(3) Adaptation of wavelet transform to Intra picture.
(4) Dynamic Vector Quantization.

(5) Super Resolution Plane Prediction.
(6) Fractal Coding

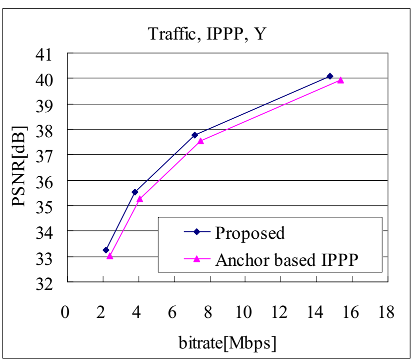
(a) Traffic (Class A)

(b) Kimono1 (Class B)

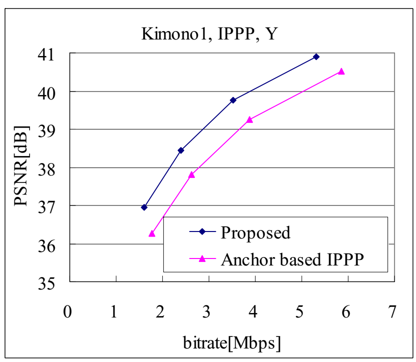
	BD PSNR [dB]	BD RATE [%]
Traffic	0.42	-12.4
Kimono1	0.89	-29.0

(c) Improvement of PSNR and Bitrate

Fig 58 Test Sequences and Coding performance



(a) Class A (Traffic)



(b) Class B (Kimono1)

Fig. 59 improvement of coding efficiency

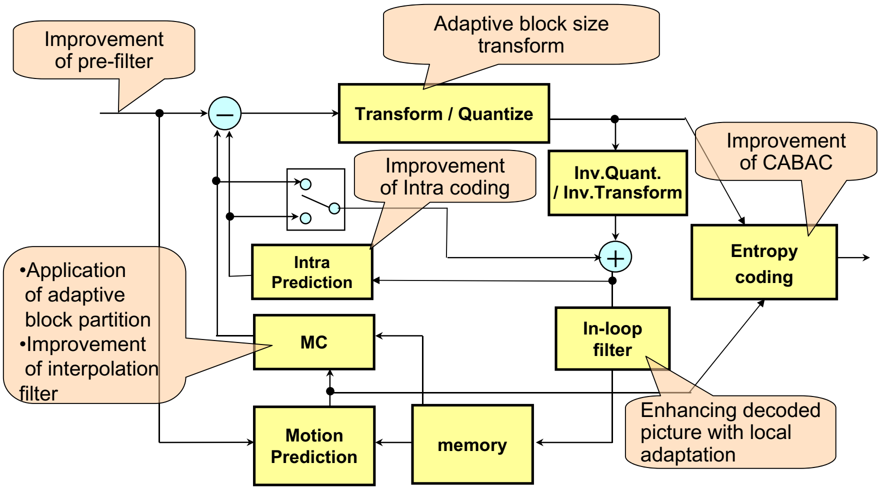


Fig. 60 Trials of coding performance improvement

7 Realization of UHD Service and System

7.1 Realization of UHD Service

UHDTV service will be realized when the next generation video coding technology is accompanied with the progress of the circumference technologies such as advanced sensor, display, storage and transmission infrastructure. UHD video signal compressed by a new video coding standard, which can realize twice as much performance as that of AVC/H.264, will be transmitted and delivered through wired (NGN) or wireless (3.9/4G. wireless LAN, satellite) lines with high speed and ubiquitous. We will be able to enjoy video with high quality and reality by means of large screen TV with over 100 inches diagonal and also to enjoy video with clearness like a gravure picture by means of organic EL display equipped with power saving and flexibility [32].

7.2 Another Possibility of Video Coding Technology in Future

Rather than extensions of the current system, high coding performance and new function for next generation video coding may also be derived from different approaches, e.g., Intelligent Coding as below:

(1) Intelligent Coding advocated by Harashima et al. in the 1980s has recently attracted attention as an alternative path for future video coding.
(2) In this technology, a common model is shared by encoder and decoder, and the model in the decoder is modified synchronously according to the information detected and transmitted by the encoder.

(3) Considering the recent progress of computer, network and image-processing technology, this method could become an interesting candidate and an important video coding technology in the future.

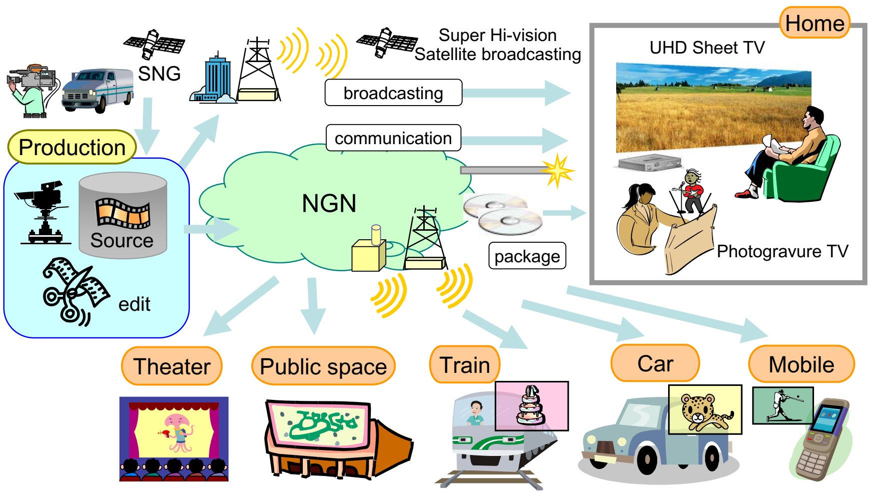


Fig. 61 Next generation video delivery through the Internet and broadcast service

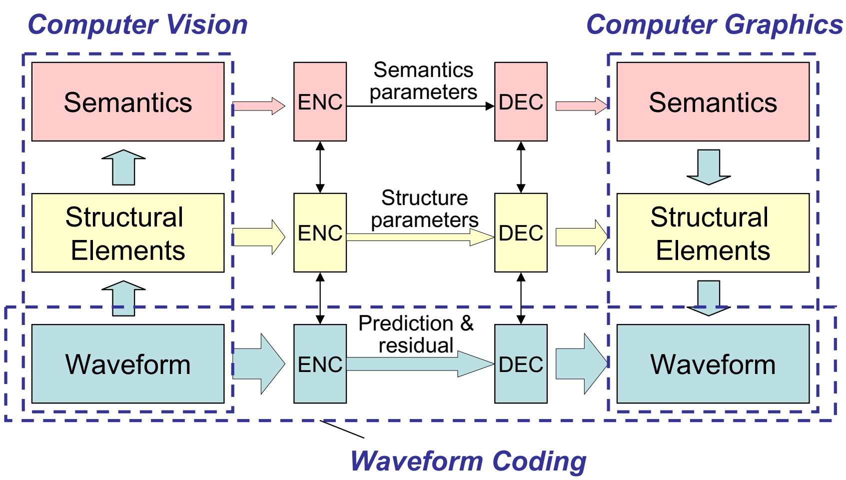


Fig. 62 Integration of graphics and computer vision to video coding [13]

Now, the performance of GPU has been markedly improved and the drawing of 8K picture can be performed on PC in real time. Therefore, a hybrid video coding method is also considered to combine Intelligent Coding and conventional coding to encode, decode and draw pictures using a high speed graphic PC. The future of video technology will spread infinitely if 3-D video technology and Free-viewpoint television technology are combined with Intelligent Coding and conventional coding technologies.

8 Summary

In 1994, MPEG-2 was standardized targeting digital TV services including HDTV by merging MPEG-3 standardization activity, then MPEG-4 standard was developed for video transmission through mobile network. One of the main goals of MPEG-4 AVC/H.264 standardization was improving video compression efficiency by two times compared to MPEG-2 to realize internet HDTV broadcasting.

The development of a new video coding standard supporting up to UHD video such as 4K cinema and 8K Super Hi-Vision is now ready to start as a collaborative work of ISO/IEC and ITU-T. It will not be far off when people can enjoy the 8K resolution video experience through the sheet type organic EL display at home.

References

1. ITU-T: Video Codec for Audiovisual Services at p×64 kbit/s, Recommendation H.261 Version 1 (1990)
2. ISO/IEC: Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s - Part 2: Video, 11172-2 (1993)
3. ISO/IEC: Generic Coding of Moving Pictures and Associated Audio Information - Part 2: Video, 13818-2 (1994)
4. ITU-T: Information technology - Generic coding of moving pictures and audio information: Video, Recommendation H.262 (1995)
5. ITU-T: Video Coding for Low Bit Rate Communication, Recommendation H.263 version 2 (1998)
6. ISO/IEC: Information technology – Coding of audio-visual objects - Part 2: Visual, 14496-2 (2004)
7. ISO/IEC: Information technology, Coding of audio-visual objects - Part 10: Advanced video coding, 14496-10 (2003)
8. ITU-T: Advanced Video Coding for generic audiovisual services, Recommendation H.264 (2003)
9. ISO/IEC: New Profiles for Professional Applications. 14496-10:2005/FDAM2 (2007)
10. Swartz, C.S. (ed.): Understanding Digital Cinema, A professional Handbook. Focal Press, USA (2005)
11. ISO/IEC: Call for Evidence on High-Performance Video Coding (HVC), JTC1/SC29/WG11 N10553 (2009)
12. Shannon, C.E.: A Mathematical Theory of Communication. Bell System Technical Journal 27, 379–423, 623–656 (1948)
13. Murakami, T.: Future of Video Coding Technology. The MPEG 20th Year Anniversary Commemoration Event, Tokyo (2008), http://www.itscj.ipsj.or.jp/forum/forum2008MPEG20.html
14. Oppenheim, A.V., Schafer, R.W.: Digital Signal Processing. Prentice Hall, USA (1975)

15. Shi, Y.Q., Sun, H.: Image and Video Compression for Multimedia Engineering. CRC press, USA (2000)
16. Murakami, T., Itoh, A., Asai, K.: Dynamic Multistage Vector Quantization of Images. Electronics and Communication in Japan 69(3), 93–101 (1986)
17. Gersho, A., Gray, R.: Vector Quantization and Signal Compression. Kluwer Academic Publishers, Dordrecht (1992)
18. Murakami, T.: International Standardization and Future Trend of Video Coding Technology. In: IEEE SSCS Kansai Chapter Technical Seminar (2006)
19. Richardson, I.E.G.: H.264 and MPEG-4 Video Compression. Wiley, England (2003)
20. ITU-R: Parameter values for an expanded hierarchy of LSDI image formats for production and international programme exchange, BT.1769 (2006)
21. Ishida, J., Suzuki, Y.: Study on the design of guardrails at Hoheikyo dam. Hokkaido Civil Engineering Research Institute monthly report Japan 588, 7–18 (2002)
22. Yamada, Y., Sekiguchi, S., Yamagishi, S., Kato, Y.: Standardization Trend of High Quality Video Coding Technology. Mitsubishi Electric technical report 82(12), 7–10 (2008)
23. SMPTE/TC-32NF-30: Proposal for 20 Gb/s Optical Interface (2009)
24. Takahashi, H.: Review of the CMOS Image Sensor Pixel Shrinking Technology. The Journal of the Institute of Image Information and Television Engineers Japan 60(3), 295–298 (2006)
25. Kubota, K.: Toward broadcasting with super-reality. In: URCF symposium, Tokyo, Japan (2009)
26. Numazawa, J.: Status and Future of Image Information Storage. The Journal of the Institute of Image Information and Television Engineers Japan 60(1), 2–5 (2006)
27. SD Association: Bridging Your Digital Lifestyle, SD Association pamphlet (2009)
28. Ishida, O.: 40/100 Gigabit Ethernet Technologies. The Journal of the Institute of Electronics, Information and Communication Engineers Japan 92(9), 782–790 (2009)
29. ISO/IEC JTC 1/SC 29/WG 11/N10361: Vision and Requirements for High-Performance Video Coding (HVC) Codec (2009)
30. ITU-T/SG16: Report of Q6/16 Rapp.Group meeting, Annex Q06.A, VCEG-AL 01 (2009)
31. Sekiguchi, S., Yamagishi, S., Itani, Y., Asai, K., Murakami, T.: On motion block size and partition for 4:4:4 video coding. ITU-T/SG16/Q.6 VCEG-AI31 (2008)
32. Murakami, T., Yoda, F.: Present Status and Future Prospect on Digital Media Technology. Mitsubishi Electric technical report Japan 82(12), 2–6 (2008)