MPEG-2

— high-compression technologies for HDTV

Masaaki Kurozumi, Yukihiro Nishida and Eisuke Nakasu
NHK STRL

Digital video coding standards offer flexibility in their encoding techniques and enable coding efficiency improvements, in compliance with the standard, over a period of time. The MPEG-2 video coding standard [1] employs the adaptive DCT coding scheme with motion-compensated prediction. The amount of overhead information, including motion vector codes and coding modes, is often large for critical HDTV sequences at lower bitrates.

NHK's new coding method [2] – conforming to the MPEG-2 Main Profile – significantly reduces the amount of overhead information and makes digital HDTV services possible at lower bitrates, while maintaining compatibility with conventional digital broadcast receivers.

In Japan, digital terrestrial television (DTT) started in December 2003, following the launch of digital satellite television in December 2000. Today, high-quality HDTV accounts for most of the digital broadcasts in the country. More than seven million households have installed digital HD receivers. All the HDTV services use the 1080/60i format and MPEG-2 Main-Profile video coding, and their bitrates are currently around 20 Mbit/s for digital satellite broadcasting and around 14 Mbit/s for digital terrestrial broadcasting. HDTV services are associated with multimedia data broadcasting services including news, weather forecasts, games, etc. HDTV data uses a large part of the transmission capacity.

One of the main features of digital broadcasting is its *extensibility* by which new kinds of broadcasting services can be introduced while maintaining existing services. Broadcasters want to enrich multimedia broadcasting services or to introduce additional TV programmes or new types of services. To enable them to do so, the HDTV bitrate will have to be significantly reduced without incurring any degradation to HDTV quality or making the existing receivers obsolete.

MPEG-2 coding scheme for interlaced video

The MPEG-2 video standard is a generic coding scheme that can be applied to diverse video formats with various resolutions, frame rates and scanning structures. MPEG-2 also offers efficient tools to encode interlaced video signals content-adaptively.

Picture structure and coding mode

Table 1 shows the MPEG-2 video syntax elements for prediction and DCT coding types. The choice of two types of picture structure (PS), frame and field, is decided in the picture layer. A **`frame_pred_frame_dct`** (FPFD) flag is available only for the frame structure. The combinations of PS and FPFD that can be chosen in the picture layer are as follows:

- PS = Frame picture, FPFD = ‘0’
- PS = Frame picture, FPFD = ‘1’
- PS = Field picture, FPFD = ‘0’

Table 1
Syntax elements of prediction type and DCT types in MPEG-2 video

Picture Layer		Macroblock Layer			
Picture _structure (PS)	**frame_pred_ frame_dct (FPFD)**	**macroblock_modes**		**Prediction type**	**DCT type**
Frame picture	‘0’	frame_motion_type (2-bit)	dct_type (1-bit)	Field-based	Frame DCT or Field DCT
				Frame-based	
				Dual-prime	
Frame picture	‘1’	–	–	Frame-based	Frame DCT
Field picture	‘0’	field_motion_type (2-bit)	–	Field-based	Field DCT
				16x8 MC	
				Dual-prime	

In the frame picture structure, prediction and DCT per macroblock are adaptively selected as indicated by **`frame_motion_type`** and **`dct_type`** in **`macroblock_modes`**. The prediction type is identified with `frame_motion_type` (2-bit) or **`field_motion_type`** (2-bit), and the DCT coding type is identified with `dct_type` (1-bit). The FPFD flag is used to reduce overhead information for progressive sequences. When `FPFD = ‘1’`, the encoding mode of the macroblock is limited to a combination of frame prediction and frame DCT, and `macroblock_modes` is not transmitted. In the field picture structure, `dct_type` is not transmitted because the DCT type is limited to only fields.

Conventional encoders usually operate with a fixed combination of `PS = frame` and `FPFD = ‘0’` (yellow-coloured rows in *Table 1*) for encoding interlaced sequences. In this case, 1 bit for I pictures and 3 bits for P and B pictures must be transmitted to identify the prediction and DCT types in a macroblock. The volume of such information for HDTV amounts to 0.7 Mbit/s – which becomes significantly large at lower bitrates.

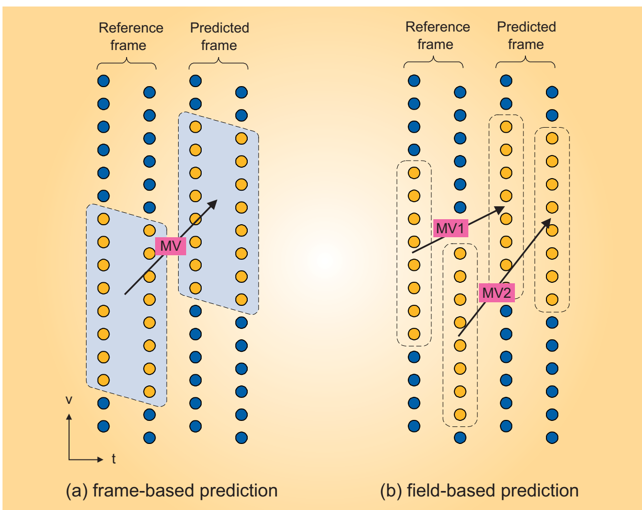

Figure 1
Adaptive prediction of frame pictures

Frame picture coding supports both frame-based and field-based predictions as shown in *Fig. 1*. Frame-based prediction works well for still or very slow motion images, whereas field-based prediction shows better performance for images with fast movement. Field-based prediction requires two motion vectors per macroblock and consumes a larger amount of motion vector codes compared with frame-based prediction.

Fig. 2 shows two DCT types of frame pictures. The frame and field DCT modes are adapted for each macroblock according to the texture and motion of the scene.

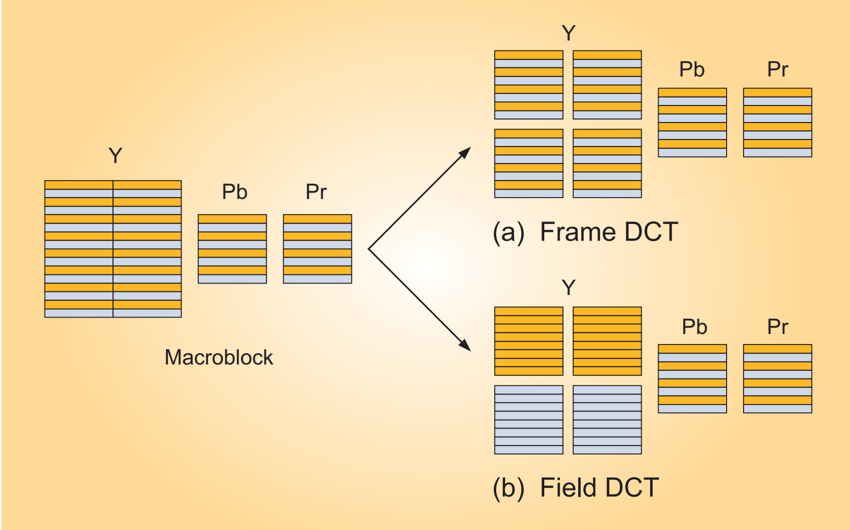

Figure 2
Adaptive DCT coding of frame pictures

Problems in coding at lower bitrates

The MPEG-2 video-coded bitstream is classified syntactically into coding modes, motion vectors and DCT coefficients. *Figs 3* and *4* show the bitstream statistics for coding-mode information (Header), motion vector (MV), and DCT coefficient code (DCT) when using a conventional MEPG-2 encoder that employs the frame picture structure (PS = frame) and FPFD = ‘0’. Since the motion vector codes account for most of the coding bits for HDTV sequences with large motions irrespective of the target bitrates, sufficient coding bits cannot be assigned to the DCT coefficients, resulting in picture-quality degradation at the lower bitrate.

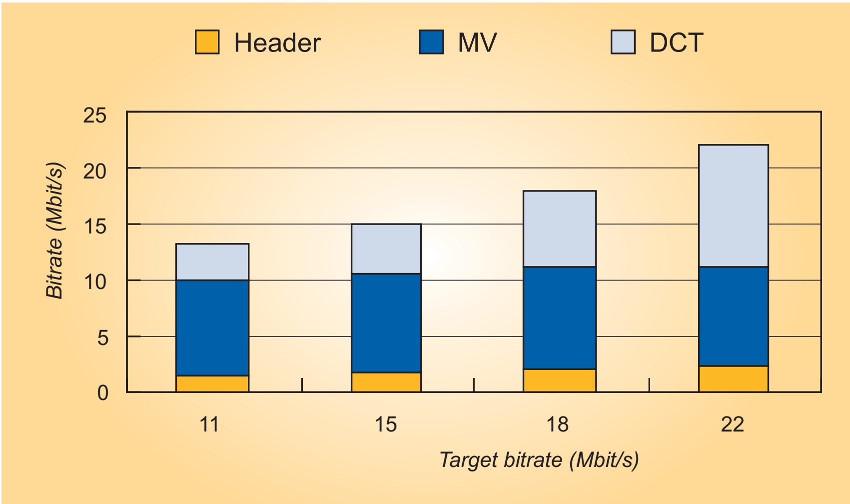

Figure 3
Overall bitstream statistics of the conventional method (“Whale Show”, MP@HL)

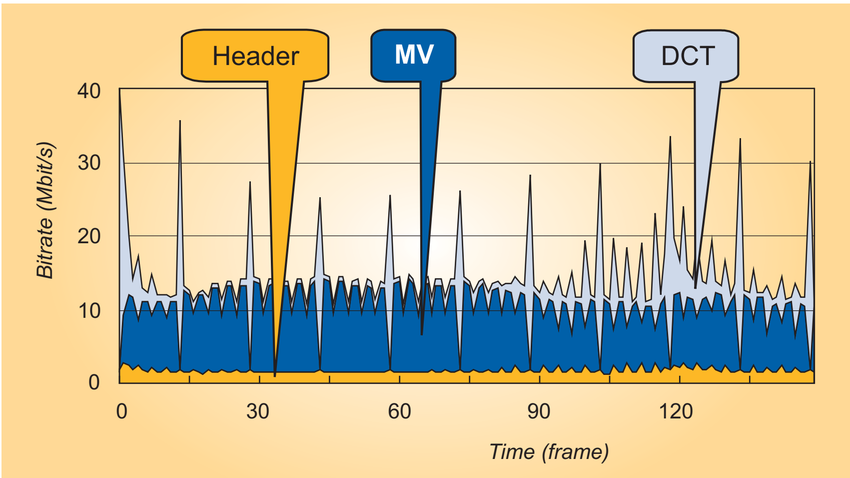

Figure 4
Frame-by-frame bitstream statistics of the conventional method (“Whale Show”, MP@HL at 15 Mbit/s)

Problems in coding of chrominance components in interlaced video

There are two major causes of colour degradation in fast-moving pictures in interlaced video:

- incorrect predictions of chrominance samples, and;

- non-adaptive DCT mode for chrominance components.

Most existing encoders use only luminance samples for motion estimation because of their simple implementation. However, doing so sometimes causes remarkable colour degradation. Moreover, in the frame-based prediction of 4:2:0 interlaced video, the prediction of chrominance samples in two fields may use opposite-parity and non-optimum temporal displacements as shown in *Fig. 5*. Such inefficient prediction will increase prediction errors for the chrominance components and will thus cause chrominance degradation.

Regarding the DCT mode, as shown in *Fig. 2*, although both frame and field DCT modes are available for the luminance components, only the frame DCT is available for the chrominance components. Therefore, the frame DCT is used to code the chrominance components even for fast-moving pictures for which the field DCT mode is more appropriate. This causes large-valued coefficients in vertical high-frequency components for interlaced images and significant chrominance degradations.

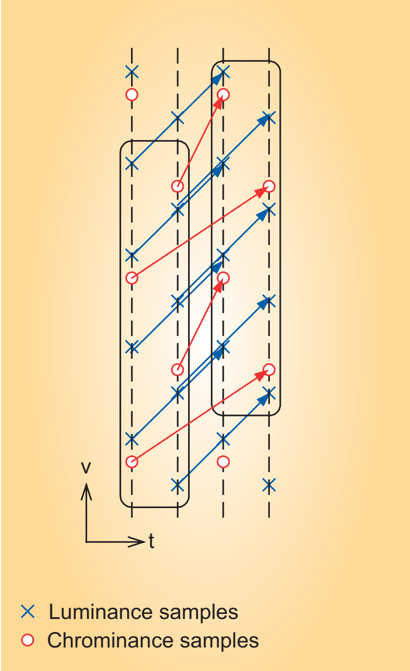

Figure 5
Example of reference samples in frame-based prediction

New coding method to solve the problems

To solve the problems of lower bitrate coding, it is essential to reduce overhead information including motion vector codes and adaptive coding modes. The problems in coding of chrominance components can be solved by content-adaptively using the field picture and by using the chrominance components for the motion estimation. Taking into account these diagnoses, our new coding method employs two schemes:

1) adaptive selection of the picture structure with a constrained coding mode, and;
2) optimized selection of the motion vectors and macroblock coding modes in MPEG-2 syntax.

Adaptive selection of picture structure

The picture structure is adaptively selected as frame or field according to the picture content. The volume of code bits will increase when high-frequency elements are included in the DCT coefficients. The proposed method determines the adequate picture structure for each picture by comparing evaluation values based on the vertical frequency energy.

The following combinations of PS and FPFD (green-coloured rows in *Table 1*) are used to reduce the amount of mode information for prediction and DCT types in a frame picture:

- PS = Frame picture, FPFD = ‘1’
- PS = Field picture, FPFD = ‘0’

In the case of PS = frame and FPFD = ‘1’, the picture consists of frame DCT blocks only, whereas in the case of PS = field and FPFD = ‘0’, the picture consists of field DCT blocks only.

Optimized selection of the motion vectors and macroblock coding modes

The new method introduces a selection rule so as to maximize the coding efficiency, which is a function of prediction error and bit-amount of the motion vector and macroblock mode. In each macroblock, the motion vector and macroblock mode are selected such that the prediction error is smaller and the assignable bit-amount for the DCT coefficients is larger. For every possible inter-macro-

Abbreviations

DCT	Discrete Cosine Transform	**MP@HL**	(MPEG-2) Main Profile at High Level
DTT	Digital Terrestrial Television	**MV**	Motion Vector
FPFD	(MPEG) frame_pred_frame_dct	**PS**	(MPEG) Picture Structure
HDTV	High-Definition Television	**PSNR**	Peak Signal-to-Noise Ratio

block mode, the ratio of the assignable bit-amount for the DCT coefficients to the mean absolute error of motion compensated prediction is evaluated while searching for the motion vector, and the motion vector that gives the maximum value is selected for each inter-macroblock mode. The inter-macroblock mode represents combinations of the direction of the prediction (forward, backward and bi-directional) and prediction type (frame, field, 16x8MC and dual-prime). The assignable bit-amount for the DCT coefficients is calculated by subtracting the bit-amounts of the macroblock mode and the motion vector from the target bit-amount for the macroblock.

The macroblock mode that gives the maximum coding efficiency is selected from among all possible macroblock modes. In the intra macroblocks, since the quantization value of the intra DC coefficient is fixed within a picture, the coding efficiency is calculated as the ratio of the assignable bit-amount for the AC coefficients to the variance of input pixels in the block. In the inter macroblocks, the coding efficiency is calculated as the ratio of the assignable bit-amount for all the DCT coefficients to the mean square error of motion compensated prediction.

The above-mentioned processes to determine the optimum set of motion vectors and macroblock mode use both the luminance and chrominance components.

Experimental results

The performance of the new method was evaluated in an encoding simulation. A comparison was also made with a conventional MPEG-2 encoder which used PS = frame and FPFD = ‘0’. *Table 2* shows the common encoding parameters. The methods used the same bitrate control. The same target bit-amount for a macroblock was used within a picture based on the bitrate control of the TM5 method [3] .

Table 2
Test conditions

Image format	1920 × 1080 /59.94i
Chroma format	4:2:0
GoP structure	N=15, M=3
Intra DC precision	8 bits
Quantizer scale type	Non-linear
Motion vector search area (horizontal × vertical)	±63.5 × ±31.5 per frame interval
Test sequences	ITU-R HDTV test materials; "European Market (Euro)", "Whale Show (Whale)" and "Green Leaves (Green)"

Figs 6 and *7* illustrate the bitstream statistics of the new method. Comparing *Figs 6* and *7* with *Figs 3* and *4*, it is apparent that the new method greatly reduces the volume of motion vector codes and increases the number of bits assigned to DCT coefficients. *Fig. 8* illustrates the PSNR improvement

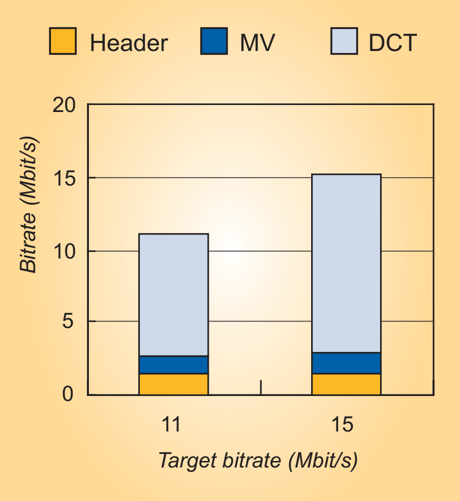

Figure 6
Overall bitstream statistics of the new method ("Whale Show", MP@HL)

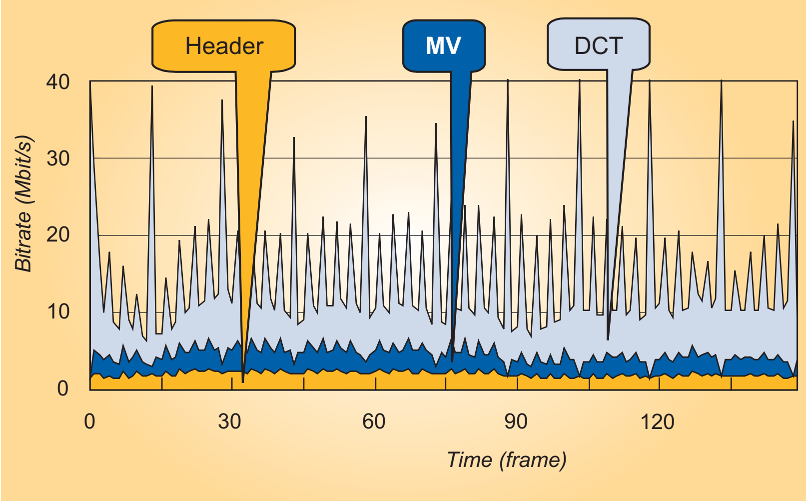

Figure 7
Frame by frame bitstream statistics of the new method ("Whale Show", MP@HL at 15 Mbit/s)

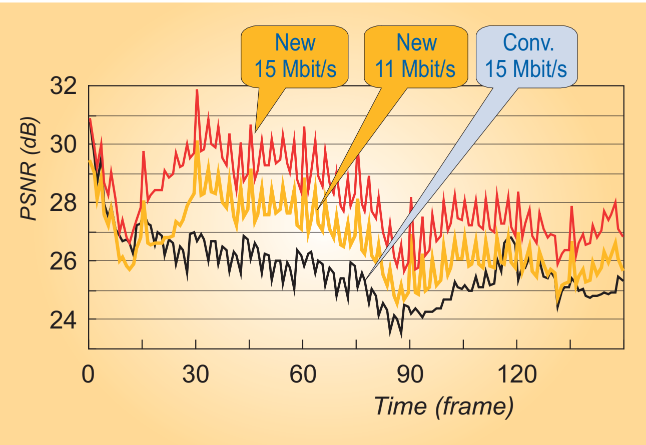

Figure 8
PSNR improvement of the new method ("Whale Show")

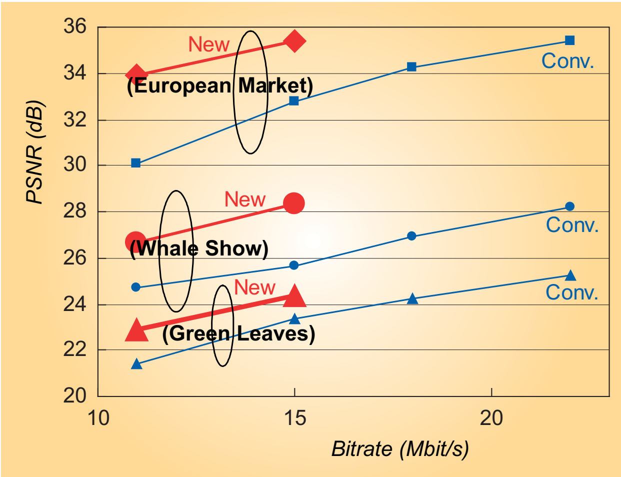

Figure 9
PSNR performance of the new method

of the new method in comparison with a conventional encoder's PSNR. *Fig. 9* compares the PSNRs of the new method and those of a conventional encoder on the three HDTV test materials coded at 11, 15, 18 and 22 Mbit/s. The PSNR improvement at 11 and 15 Mbit/s was 1 to 4 dB. The new method's PSNR at 15 Mbit/s and 11 Mbit/s was comparable to the conventional encoder's PSNR at 18 to 22 Mbit/s and 15 to 18 Mbit/s.

Fig. 10 (on the next page) shows cropped portions of HDTV pictures coded at 15 Mbit/s by a conventional encoder and by the new method. The new method reduces the block distortion and significantly improves the subjective picture quality. It preserves spatial detail in all scenes, and the improvement of picture quality is especially notable in fast-moving scenes.

Conclusions

The MPEG-2 video coding standard provides several content-adaptive coding algorithms. If a video encoder sufficiently exploits the video-coding algorithms as defined in the standard, coding efficiency can be improved while keeping compatibility with legacy receivers. The new coding method reported here efficiently reduces the overhead information and improves the picture quality. The key technologies of this method include ***adaptive selection of picture structure*** and ***optimized selection of coding mode***, which both conform to the MPEG-2 Main Profile.

(a) Conventional MPEG-2 encoder

(b) New method

Figure 10
Coded picture of the sequence "Whale Show" (MP@HL at 15Mbit/s)
Note: the image size shown is 640x480 pixels, cropped from 1920x1080 pixels

Masaaki Kurozumi received a BE degree in Electronics from Doshisha University, Japan, in 1993, and an ME degree in Information Processing from the Nara Institute of Science and Technology in 1995. In the same year, he joined NHK (Japan Broadcasting Corporation) and worked as a broadcasting engineer at the Osaka Broadcasting Station. Since 1997, he has been with NHK Science and Technical Research Laboratories and is engaged in research on picture coding.

Yukihiro Nishida is an Associate Director, Visual Information Technologies, at NHK Science and Technical Research Laboratories. Since graduating from Keio University, Japan, with a Master Degree in Electrical Engineering, he has worked for NHK (Japan Broadcasting Corporation) and has been involved in research and development of video coding, quality measurement, digital broadcasting, HDTV, etc. He has frequently participated in standardization activities in the area of broadcasting, and is a Vice-Chairman of Working Party 6A in the ITU-R. He participates in EBU project group B/TQE (Television Quality Evolution). In 1992, he spent six months at RAI Research Centre in Italy, working on video coding technology.

Eisuke Nakasu received BE and ME degrees in Electrical Engineering from Keio University, Japan, in 1980 and 1982, respectively. He joined NHK (Japan Broadcasting Corporation) in 1982 and is in charge of research and development for digital broadcasting systems, video coding and quality evaluation. He is a Senior Research Engineer, Visual Information Technologies, at NHK Science and Technical Research Laboratories. He participated in the MPEG-2 standardization activities and contributed to the development and standardization of the digital satellite broadcasting and digital terrestrial broadcasting systems in Japan. Currently he is a chairman of the Study Group on Quality Assessment Methods within ARIB, Japan.

Although new video coding standards such as MPEG-4 AVC/H.264 have emerged and are expected to be introduced in new digital media including broadcasting, MPEG-2 will survive for some time to come yet – especially in situations where MPEG-2 is already in use. The new coding method for high-compression MPEG-2 will be a good option for broadcasters and manufacturers who wish to enhance the picture quality or reduce the channel capacity.

References

[1] ISO/IEC 13818-2:2000 | ITU-T Rec. H.262 (2000): **Information technology – Generic coding of moving pictures and associated audio information: Video**
http://www.itu.int/ITU-T/publications/recs.html

[2] M. Kurozumi et al.: **Lower bitrate HDTV MPEG-2 coding by adaptive selection of picture structure**
Proc. of PCS 2003, pp. 317 to 320 (2003)
http://www-doc.irisa.fr/cgi-bin/bibvisu?OCLC_W3=34265

[3] ISO/IEC JTC1/SC29/WG11/N0400 (1993): **"Test Model 5"**