REPORT ITU-R BT.2017

STEREOSCOPIC TELEVISION MPEG-2 MULTI-VIEW PROFILE

(1998)

1 Overview of MPEG-2 multi-view profile (MVP)

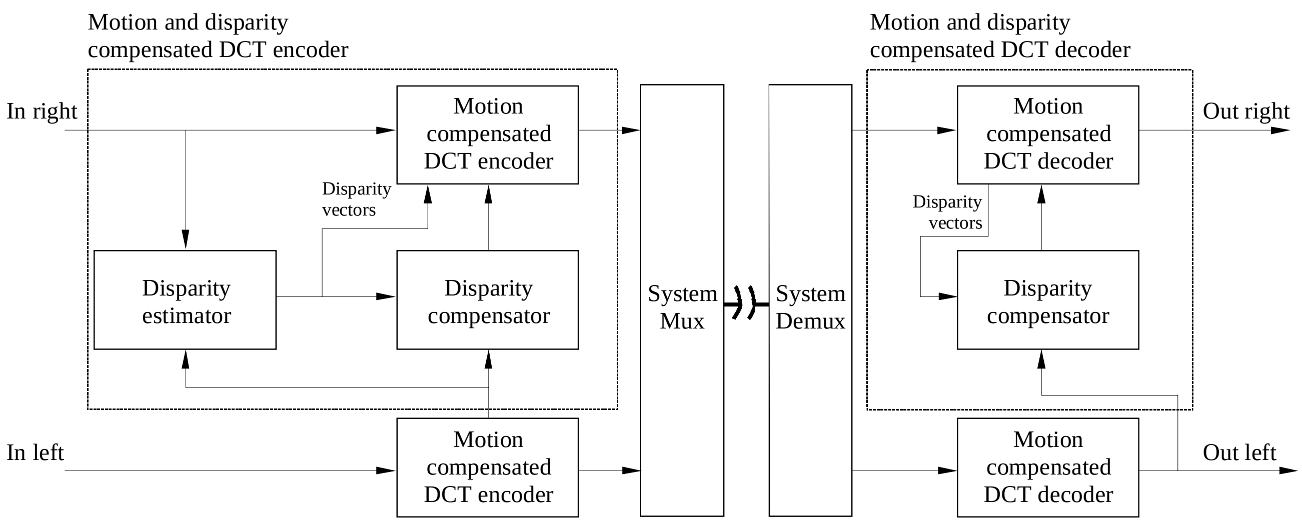
The extension of the MPEG-2 video standard (ITU-T Recommendation H.262 | ISO/CEI 13818-2: Information technology – Generic coding of moving pictures and associated audio information: Video) for multi-view applications (e.g. used for stereoscopic video) has been promoted to a final International Standard at the ISO/IEC JTC 1/SC 29/WG 11 meeting in September 1996 (Amendment 3, WG 11 N1366) the multi-view profile (MVP) is envisioned to be a profile appropriate for applications that require multiple viewpoints within the context the MPEG-2 video standard. MVP supports stereoscopic pictures as its source images for a wide range of picture resolution and quality as requested by the applications to be used.

1.1 Coding scheme for MVP

A block diagram of the codec reference model for the MVP is shown in Fig. 1. Its main features are a monoscopic coding in its base layer for compatibility and a hybrid prediction of motion and disparity for compression efficiency. Temporal scalability tools are used for coding an enhancement layer.

FIGURE 1

The codec reference model for the MVP



DCT : discret cosine transform

Rap 2017-01

A configuration of prediction modes are shown in Fig. 2. A monoscopic coding with the same tools as main profile (MP), including the ISO/IEC 11172-2 Standard is applied to the base layer. A base layer of MVP is assigned to a left view and an enhancement layer is assigned to a right view. An enhancement layer is coded using temporal scalability tools and a hybrid prediction of motions and disparity can be utilized in the enhanced layer. It foresees higher compression of the right view of stereoscopic video by exploiting the similarity between the left and right views.

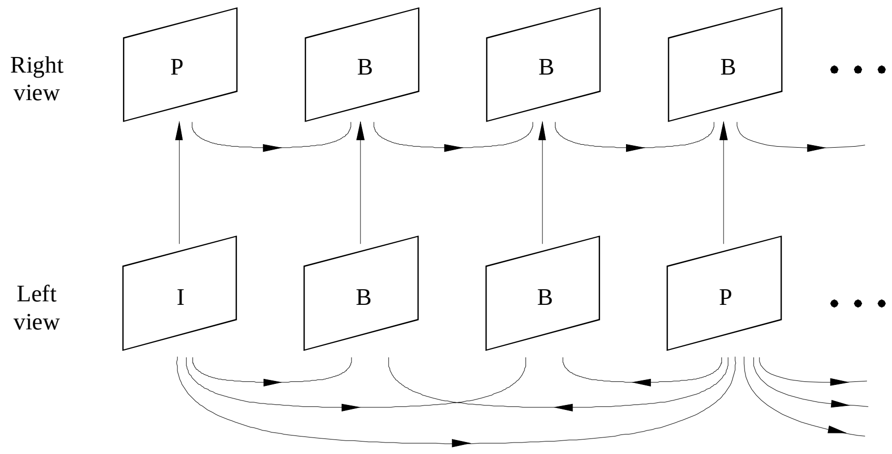
MVP, one of the scalable profiles in terms of multiple viewpoint layers, has the same type of compatibility features; other scalable profiles have such compatibility with MP. For example:

– decoders compliant to MVP at a certain level are capable of decoding the bitstreams compliant to MP at the corresponding level (i.e. forward compatibility)

– decoders compliant to MP at a certain level are capable of decoding the bitstream in the base layer of MVP (i.e. backward compatibility).

FIGURE 2

Prediction configuration example with M = 3 coding of left-view, right-view frame picture coded using disparity prediction with respect to left-view and motion prediction with respect to itself



Rap 2017-02

1.2 Parameter values for MVP

The levels for the MVP are high, high-1440, main and low. Temporal scalability involves two layers, a base layer and an enhancement layer. Both the enhancement and base layers have the same spatial resolution at the same frame rate. Tables 1 to 4 present bounds on sampling rates, luminance pel rates, bit rates and buffer sizes for the MVP.

TABLE 1

Upper bounds for sampling density

Level	Spatial resolution layer		Profile
			Multiview
High	Enhancement (right view)	Samples/line Lines/frame Frames/s	1 920 1152 60
	Lower (left view)	Samples/line Lines/frame Frames/s	1 920 1152 60
High-1440	Enhancement (right view)	Samples/line Lines/frame Frames/s	1 440 1152 60
	Lower (left view)	Samples/line Lines/frame Frames/s	1 440 1152 60
Main	Enhancement (right view)	Samples/line Lines/frame Frames/s	720 576 30
	Lower (left view)	Samples/line Lines/frame Frames/s	720 576 30
Low	Enhancement (right view)	Samples/line Lines/frame Frames/s	352 288 30
	Lower (left view)	Samples/line Lines/frame Frames/s	352 288 30

TABLE 2

Upper bounds for luminance sample rate (samples/s)

Level	Spatial resolution layer	Profile
		Multiview
High	Enhancement (right view)	62 668 800
	Lower (left view)	62 668 800
High-1440	Enhancement (right view)	47 001 600
	Lower (left view)	47 001 600
Main	Enhancement (right view)	10 368 000
	Lower (left view)	10 368 000
Low	Enhancement (right view)	3 041 280
	Lower (left view)	3 041 280

TABLE 3

Upper bounds for bit rates (Mbit/s)

Level	Profile
	Multiview
High	130 both layers
	80 base layer
High-1440	100 both layers
	60 base layer
Main	25 both layers
	15 base layer
Low	8 both layers
	4 base layer

TABLE 4

Buffer size requirements (bits)

Level	Layer	Profile
		Multiview
High	Enhancement	15 898 480
	Base	9 787 248
High-1440	Enhancement	12 222 464
	Base	7 340 032
Main	Enhancement	3 047 424
	Base	1 835 008
Low	Enhancement	950 272
	Base	475 136

1.3 Camera parameter extension

An extension for camera information has been introduced in MVP. The extension specifies the height of image device, the focal length, the F-number, the vertical angle of the field of view, the position and the direction of the camera, and upper direction of the camera.

2 Assessment tests for MVP

The verification tests for the MVP were carried out at three different test sites located in Japan, Germany and Canada. The results of tests were presented at the WG 11 Chicago meeting (WG 11 N1373, September 1996. Test and video subgroup "Results of MPEG-2 multiview profile verification test"). The results of the different test sites are consistent with each other and show that in general, at the tested bit rates, the observers judged that the MPEG-2 multi-view profile coding scheme did not introduce annoying coding artifacts.

2.1 Test method

The double stimulus impairment scale method (variant II) in Recommendation ITU-R BT.500 was applied. Instead of the discrete scale recommended by ITU-R a continuous scale was used in order to obtain more precise evaluations.

2.2 Test conditions

The test sequences generated during bit stream exchange were used. An overview of the test conditions is provided in Table 5. Different display systems were used at each test site.

TABLE 5

Overview of the subjective test conditions

Sequences	"Street organ", "Flower pot", "Trapeze" (525/60) "Fun fair" (625/50)
Algorithms and bit-rates (left/right view)	MVP@ML: 6/3 Mbit/s, 9/4 Mbit/s Simulcast of MP@ML: 4.5/4.5 Mbit/s, 6.5/6.5 Mbit/s Simulcast of MP@ML as lower anchor: 2.5/2.5Mbit/s (for "Street organ", "Fun fair"), 1.5/1.5 Mbit/s (for "Flower pot", "Trapeze") Original/original as upper anchor
Test method	The double-stimulus impairment scale method (variant II) described in Recommendation ITU-R BT.500, with a continuous scale
Stereoscopic display system (picture size, viewing distance)	HHI: two-mirror display system (19 cm × 14 cm, 5 *H*) CRC: time sequential display and LCD shutter eyeglasses (40.6 cm × 30.5 cm, 4 *H*) NHK: LCD high definition television (HDTV) projectors and polarizing eyeglasses (82 cm × 57 cm, 5 *H*)
Observers	HHI: 24 non-expert viewers CRC: 18 non-expert viewers NHK: 19 non-expert viewers (an observer was rejected by screening based on Recomnebdation TU-R BT.500)

HHI: Heinrich-Hertz-Institut für Nachrichtentechnik (Germany)

CRC: Communications Research Center (Canada)

NHK: Nippon Hoso (Kyokai (Japan)

2.3 Results of subjective assessment tests

The mean scores and the 95% confidence intervals of the means were calculated for each test condition. The test results of HHI, CRC and NHK are provided in Table 6 and Fig. 3. HHI1 and HHI2 are results obtained at HHI on two different parts of the same sequence. HHI could not test the whole sequences because of limitations of display memory size.

TABLE 6

Mean scores and the 95% confidence intervals

a) Sequence: Street organ

	Source	MVP (9/4 Mbit/s)	MP × 2 (6.5/6.5 Mbit/s)	MVP (6/3 Mbit/s)	MP × 2 (4.5/4.5 Mbit/s)	Lower anchor
NHK	4.71 ±0.17	4.18 ±0.27	4.40 ±0.26	4.06 ±0.39	3.51 ±0.32	1.74 ±0.33
CRC	4.24 ±0.37	4.19 ±0.33	4.33 ±0.29	4.27 ±0.34	4.07 ±0.35	2.19 ±0.35
HHI1	4.89 ±0.12	4.55 ±0.21	4.58 ±0.22	4.23 ±0.26	3.63 ±0.35	1.30 ±0.19
HHI2	4.86 ±0.13	4.68 ±0.19	4.85 ±0.13	4.44 ±0.24	4.24 ±0.32	1.80 ±0.23

b) Sequence: Flower pot

	Source	MVP (9/4 Mbit/s)	MP × 2 (6.5/6.5 Mbit/s)	MVP (6/3 Mbit/s)	MP × 2 (4.5/4.5 Mbit/s)	Lower anchor
NHK	4.79 ±0.16	4.03 ±0.44	4.28 ±0.25	4.07 ±0.33	4.13 ±0.37	2.28 ±0.32
CRC	4.53 ±0.14	4.57 ±0.20	4.45 ±0.22	4.40 ±0.20	4.40 ±0.21	2.70 ±0.34
HHI1	4.81 ±0.19	4.49 ±0.25	4.52 ±0.26	4.33 ±0.24	4.46 ±0.23	1.96 ±0.25
HHI2	4.83 ±0.14	4.48 ±0.21	4.33 ±0.22	4.08 ±0.26	4.16 ±0.25	1.69 ±0.24

c) Sequence: Trapeze

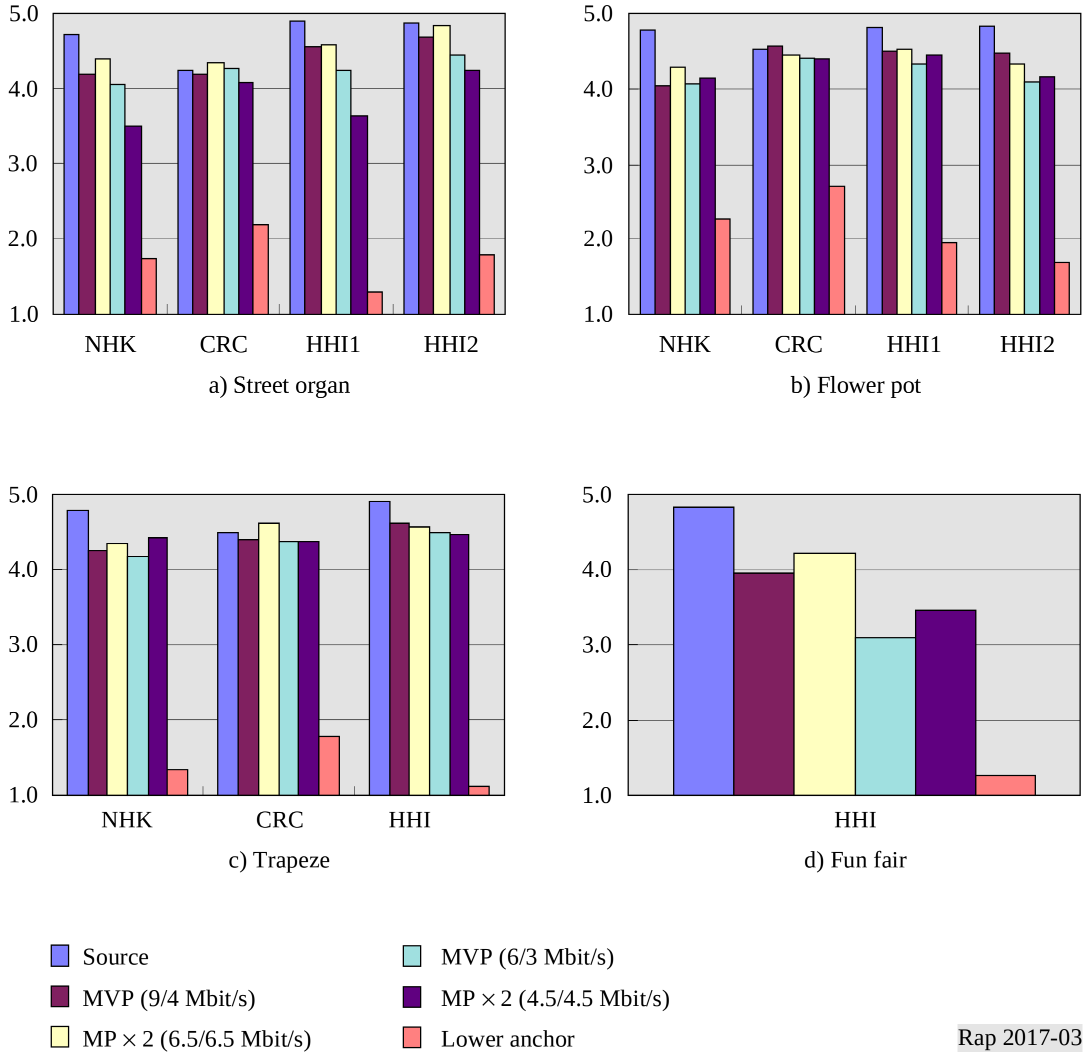
	Source	MVP (9/4 Mbit/s)	MP × 2 (6.5/6.5 Mbit/s)	MVP (6/3 Mbit/s)	MP × 2 (4.5/4.5 Mbit/s)	Lower anchor
NHK	4.77 ±0.13	4.24 ±0.25	4.34 ±0.38	4.16 ±0.24	4.41 ±0.23	1.33 ±0.18
CRC	4.48 ±0.22	4.38 ±0.24	4.62 ±0.14	4.37 ±0.23	4.36 ±0.24	1.78 ±0.31
HHI1	4.90 ±0.11	4.60 ±0.19	4.55 ±0.25	4.48 ±0.27	4.46 ±0.28	1.13 ±0.14

d) Sequence: Fun fair

	Source	MVP (9/4 Mbit/s)	MP × 2 (6.5/6.5 Mbit/s)	MVP (6/3 Mbit/s)	MP × 2 (4.5/4.5 Mbit/s)	Lower anchor
HHI1	4.83 ±0.14	3.96 ±0.32	4.23 ±0.29	3.10 ±0.35	3.46 ±0.27	1.27 ±0.20

FIGURE 3

Mean scores of subjective assessment



Some aspects of the results of this test seem worth being mentioned:

– Within each of the four sequences, the mean score of the MVP sequence at a bit rate of 9/4 Mbit/s does not differ significantly from the mean score of the simulcast of MPs at a bit rate of 6.5/6.5 Mbit/s. As well, the mean score of the MVP sequence at a bit rate of 6/3 Mbit/s does not differ significantly from the mean score of the simulcast of MPs at a bit rate of 4.5/4.5 Mbit/s, except the one pair of the sequence "Street organ." For "Street organ," the quality of MVP is superior to that of simulcast of MPs. These results show that differences in subjective evaluation between MVP and simulcast of MPs are very small at higher bit rate for pictures with slight motions ("Flower pot" and "Trapeze") and/or with significant luminance difference between left and right views ("Fun fair").

– "Fun fair" is the scene with the most differing mean scores. In this scene the most movement (changes of the image content to the next frame) could be observed in comparison with the other scenes. Especially in "Fun fair", the moving objects cover most of the image.

3 Future work on stereoscopic television

Progress made up to now has provided evidence that stereoscopic television is technically feasible. The recently approved MPEG multi-view profile provides a basis for coding and compression of stereoscopic video sequences. The quality assessment tests carried out also have brought evidence that, within the limits of the test parameters chosen, subjectively perceived satisfactory picture quality can be achieved. Nevertheless, still many issues remain unanswered. Some of the issues that require further information are as follows:

3.1 Requirement

– It will be desirable that any future stereoscopic television system is compatible with the currently emerging monoscopic digital television systems, and additional bit rate should be as small as possible.

– The quality of the monoscopic main picture that may be viewed on a monoscopic television display should be as close to that of the quality of a monoscopic picture using the entire channel capacity.

3.2 Required information for both standard definition television (SDTV) and HDTV

– The degree of asymmetric bit-rate allocation that is possible to the left- and right-view pictures for a stereoscopic video sequence to achieve minimal quality degradation for the base level picture.

– The effect of asymmetric bit rate allocation to the left- and right-view pictures on the subjectively perceived coding and compression artefacts and overall quality of the stereoscopic video sequence.

– The factors that may lead to viewer fatigue; and mitigating measures that could reduce or eliminate such fatigue.

– The required bit rate range to achieve subjectively perceived satisfactory quality for both the stereoscopic picture as well as the monoscopic picture provided by the base level picture, through additional assessment tests with a large number of video sequences representing a wide range of programming material and for a wide bit-rate range.

– Appropriate assessment test methods for stereoscopic images.

– Coding algorithms that enable more efficient compression of stereoscopic television signals.

These studies should be carried out maintaining the liaison with WP 11B, JWP 10-11Q, and other relevant working parties and organizations.