INTERNATIONAL STANDARD

ISO/IEC 14496-2

Second edition
2001-12-01

AMENDMENT 1
2002-02-01

Information technology — Coding of audio-visual objects —

Part 2: **Visual**

AMENDMENT 1: Studio profile

Technologies de l'information — Codage des objets audiovisuels —

Partie 2: Codage visuel

AMENDEMENT 1: Profil du studio

Reference number
ISO/IEC 14496-2:2001/Amd.1:2002(E)

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 3.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this Amendment may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Amendment 1 to International Standard ISO/IEC 14496-2:2001 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

Information technology — Coding of audio-visual objects — Part 2: Visual

AMENDMENT 1: Studio profile

1) Add the following text at the end of 'Overview of the object based non scalable syntax' of 'Introduction':

"

In order to preserve the lossless quality, or to restrict the maximum bit count of block data, the block based DPCM coding can be used for ISO/IEC 14496-2:2001 Amendment 1 (Studio Profile Amendment).

"

2) Replace text in 'Coding of Shapes' of 'Introduction',

"

In natural video scenes, VOPs are generated by segmentation of the scene according to some semantic meaning. For such scenes, the shape information is thus binary (binary shape). Shape information is also referred to as alpha plane. The binary alpha plane is coded on a macroblock basis by a coder which uses the context information, motion compensation and arithmetic coding.

"

with

"

In natural video scenes, VOPs are generated by segmentation of the scene according to some semantic meaning. For such scenes, the shape information is thus binary (binary shape). Shape information is also referred to as alpha plane. The binary alpha plane is coded on a macroblock basis by a coder which uses the context information, motion compensation and arithmetic coding. For high quality applications, the uncompressed binary alpha block coding is used.

"

3) Add the following text in 'Introduction' following 'Coding of Shapes':

"

Coding interlaced video

Each frame of interlaced video consists of two fields which are separated by one field-period. This part of ISO/IEC 14496 allows either the frame to be encoded as a VOP or the two fields to be encoded as two VOPs. Frame encoding or field encoding can be adaptively selected on a frame-by-frame basis. Frame encoding is typically preferred when the video scene contains significant detail with limited motion. Field encoding, in which the second field can be predicted from the first, works better when there is fast movement.

"

4) Replace text in ‘Motion representation - macroblocks' of ‘Introduction’,

"

The choice of 16×16 blocks (referred to as macroblocks) for the motion-compensation unit is a result of the trade-off between the coding gain provided by using motion information and the overhead needed to represent it. Each macroblock can further be subdivided to 8×8 blocks for motion estimation and compensation depending on the overhead that can be afforded. In order to encode the highly active scene with higher vop rate, a Reduced Resolution VOP tool is provided. When this tool is used , the size of the macroblock used for motion compensation decoding is 32 x 32 pixels and the size of block is 16 x 16 pixels.

"

with

"

The choice of 16×16 blocks (referred to as macroblocks) for the motion-compensation unit is a result of the trade-off between the coding gain provided by using motion information and the overhead needed to represent it. Each macroblock can further be subdivided to 8×8 blocks for motion estimation and compensation depending on the overhead that can be afforded. In order to encode the highly active scene with higher vop rate, a Reduced Resolution VOP tool is provided. When this tool is used , the size of the macroblock used for motion compensation decoding is 32 x 32 pixels and the size of block is 16 x 16 pixels.

In frame encoding, the prediction from the previous reference frame can itself be either frame-based or field-based.

"

5) Replace text in ‘Chrominance formats’ of ‘Introduction’,

"

This part of ISO/IEC 14496 currently supports the 4:2:0 chrominance format.

"

with

"

This part of ISO/IEC 14496 currently supports the 4:2:0 chrominance format.

ISO/IEC 14496-2:2001 Amendment 1 also supports the 4:2:2 and 4:4:4 choromінance formats in addition.

"

6) Add the following text in ‘Introduction’ following ‘Chrominance formats’:

"

RGB color components

ISO/IEC 14496-2:2001 Amendment 1 supports coding of RGB color components. The resolution of each component shall be identical when input data is treated as RGB color components.

7) Add the following text at the end of ‘Pixel depth’ of ‘Introduction’:

"

ISO/IEC 14496-2:2001 Amendment 1 supports 8, 10 and 12 bits in luminance and chrominance or RGB planes.

"

8) Replace subclauses 3.38, 3.82, 3.107, and 3.131 with the following:

"

3.38 **component**: A matrix, block or single sample from one of the three matrices (luminance and two chrominance or green, blue and red color primaries) that make up a picture.

3.82 **frame**: A frame contains lines of spatial information of a video signal. For progressive video, these lines contain samples starting from one time instant and continuing through successive lines to the bottom of the frame. For interlaced video a frame consists of two fields, a top field and a bottom field. One of these fields will commence one field period later than the other.

3.107 **macroblock**: The four 8×8 blocks of luminance data and the two (for 4:2:0 chrominance format), four (for 4:2:2 chrominance format) or eight (for 4:4:4 chrominance format) corresponding 8×8 blocks of chrominance data coming from a 16×16 section of the luminance component of the picture. Macroblock is sometimes used to refer to the sample data and sometimes to the coded representation of the sample values and other data elements defined in the macroblock header of the syntax defined in this part of ISO/IEC 14496. The usage is clear from the context.

3.131 **picture**: Source, coded or reconstructed image data. A source or reconstructed picture consists of three rectangular matrices of N-bit numbers representing the luminance and two chrominance signals or rgb colour signals. A “coded VOP” was defined earlier. For progressive video, a picture is identical to a frame, while for interlaced video, a picture can refer to a frame, or the top field or the bottom field of the frame depending on the context.

"

9) Add the following subclauses in clause 3 and renumber the subsequent items.

"

3.6 **B-field VOP**: A field structure B-VOP.

3.7 **B-frame VOP**: A frame structure B-VOP.

3.20 **bottom field**: One of two fields that comprise a frame. Each line of a bottom field is spatially located immediately below the corresponding line of the top field.

3.33 **coded B-frame**: A B-frame VOP or a pair of B-field VOPs that is coded.

3.34 **coded frame**: A coded frame is a coded I-frame, a coded P-frame or a coded B-frame.

3.35 **coded I-frame**: An I-frame VOP or a pair of field VOPs that is coded where the first field VOP is an I-VOP and the second field VOP is an I-VOP or a P-VOP..

3.36 **coded P-frame**: A P-frame VOP or a pair of field VOPs that is coded.

3.42 **coded order**: The order in which the VOPs are transmitted and decoded. This order is not necessarily the same as the display order.

3.64 **display aspect ratio**: The ratio height/width (in spatial measurement units such as centimeters) of the intended display.

3.66 **display process**: The (non-normative) process by which reconstructed frames are displayed.

3.85 **fast forward playback**: The process of displaying a sequence, or parts of a sequence, of VOPs in display-order, faster than real-time.

3.86 **fast reverse playback**: The process of displaying a sequence, or parts of a sequence, of VOPs in the reverse of display order, faster than real-time.

3.88 **field**: For an interlaced video signal, a “field” is the assembly of alternate lines of a frame. Therefore an interlaced frame is composed of two fields, a top field and a bottom field.

3.89 **field-based prediction**: A prediction mode using only one field of the reference frame. The predicted block size is 16x16 luminance samples. Field-based prediction is not used in progressive frames.

3.90 **field period**: The reciprocal of twice the frame rate.

3.91 **field VOP; field structure VOP**: A field structure VOP is a coded VOP with vop_structure is equal to “Top field” or “Bottom field”.

3.99 **frame-based prediction**: A prediction mode using both fields of the reference frame.

3.102 **frame VOP; frame structure VOP**: A frame structure VOP is a coded VOP with vop_structure is equal to “Frame”.

3.103 **future reference frame (field)**: A future reference frame (field) is a reference frame (field) that occurs at a later time than the current VOP in display order.

3.113 **I-field VOP**: A field structure I-VOP.

3.114 **I-frame VOP**: A frame structure I-VOP.

3.147 **RGB component**: A matrix, block or single sample representing one of the three primary colours. The symbols used for the rgb signals are Green, Blue and Red.

3.148 **P-field VOP**: A field structure P-VOP.

3.149 **P-frame VOP**: A frame structure P-VOP.

3.171 **sample aspect ratio**: (abbreviated to **SAR**). This specifies the relative distance between samples. It is defined (for the purposes of this specification) as the vertical displacement of the lines of luminance samples in a frame divided by the horizontal displacement of the luminance samples. Thus its units are (metres per line) ÷ (metres per sample)

3.182 **skipped macroblock**: A macroblock for which no data is encoded.

3.192 **top field**: One of two fields that comprise a frame. Each line of a top field is spatially located immediately above the corresponding line of the bottom field.

"

10) Add the following subclause 5.2.9 after subclause 5.2.8:

"

5.2.9 Definition of next_start_code_studio() function

The next_start_code_studio() function removes any zero bit and zero byte stuffing and locates the next start code.

next_start_code_studio() {	**No. of bits**	**Mnemonic**
while (!bytealigned())		
zero_bit	1	‘0’
while (nextbits() != ‘0000 0000 0000 0000 0000 0001’)		
zero_byte	8	‘0000 0000’
}		

This function checks whether the current position is byte aligned. If it is not, zero stuffing bits are present. After that any number of zero stuffing bytes may be present before the start code. Therefore start codes are always byte aligned and may be preceded by any number of zero stuffing bits.

"

11) Replace subclause 6.1.1 with the following:

"

6.1.1 Visual object sequence

Visual object sequence is the highest syntactic structure of the coded visual bitstream.

A visual object sequence commences with a visual_object_sequence_start_code which is followed by profile_and_level_indication, and one or more visual objects coded concurrently. The visual object sequence is terminated by a visual_object_sequence_end_code.

At various points in the visual object sequence, a repeat visual_object_sequence_start_code can be inserted for coded video data. In that case, the repeat visual_object_sequence_start_code shall follow a particular VOP.

When profile_and_level_indication indicates a Studio Profile, StudioVisualObject() shall follow it.

"

12) Replace subclause 6.1.2 with the following:

"

6.1.2 Visual object

A visual object commences with a visual_object_start_code and a visual object id, which are followed by a video object, a still texture object, a mesh object, or an FBA object.

For Studio Profiles, only video object type is supported.

"

13) Replace subclause 6.1.3 with the following:

"

6.1.3 Video object

A video object commences with a video_object_start_code, and is followed by one or more video object layers.

A video object layer commences with video_object_layer_start_code which may optionally be followed by Group_of_StudioVideoObjectPlane() and then by one or more coded VOPs. The order of the coded frames in the coded bitstream is the order in which the decoder processes them, which is not necessarily the display order.

"

14) Replace subclause 6.1.3.1 with the following:

"

6.1.3.1 Progressive and interlaced sequences

This part of ISO/IEC 14496 deals with coding of both progressive and interlaced sequences.

The sequence, at the output of the decoding process, consists of a series of reconstructed VOPs separated in time and are readied for display via the compositor.

For Studio Profiles paticularly, the output of the decoding process for interlaced sequences consists of a series of reconstructed fields that are separated in time by a field period. The two fields of a frame may be coded separately (field-VOPs). Alternatively the two fields may be coded together as a frame (frame-VOPs). Both frame VOPs and field VOPs may be used in a single video sequence.

In progressive sequences each VOP in the sequence shall be a frame VOP. The sequence, at the output of the decoding process, consists of a series of reconstructed frames that are separated in time by a frame period.

"

15) Replace subclause 6.1.3.2 with the following :

"

6.1.3.2 Frame

A frame consists of three rectangular matrices of integers; a luminance matrix (Y), and two chrominance matrices (Cb and Cr).

The relationship between these Y, Cb and Cr components and the primary (analogue) Red, Green and Blue Signals (E'_R, E'_G and E'_B), the chromaticity of these primaries and the transfer characteristics of the source frame may be specified in the bitstream (or specified by some other means). This information does not affect the decoding process.

For Studio Profiles particularly, the three rectangular matrices can be the primary RGB colour matrices.

"

16) Add the following subclause in subclause 6.1.3 and renumber the subsequent items

"

6.1.3.3 Field

A field consists of every other line of samples in the three rectangular matrices of integers representing a frame.

A frame is the union of a top field and a bottom field. The top field is the field that contains the top-most line of each of the three matrices. The bottom field is the other one.

"

17) Replace subclause 6.1.3.3 with the following:

"

6.1.3.3 VOP

A reconstructed VOP is obtained by decoding a coded VOP. A coded VOP may have been derived from a progressive or interlaced frame or an interlaced field. A reconstructed VOP is either a reconstructed frame (when decoding a frame VOP), or one field of a reconstructed frame (when decoding a field VOP).

An I-frame VOP or a pair of field VOPs, where the first field VOP is an I-picture and the second field VOP is an I-VOP or a P-VOP, is called a coded I-frame.

A P-frame VOP or a pair of P-field VOPs is called a coded P-frame.

A B-frame VOP or a pair of B-field VOPs is called a coded B-frame.

A coded I-frame, a coded P-frame or a coded B-frame is called a coded frame.

6.1.1.4.1 Field VOPs

If field VOPs are used, then they shall occur in pairs (one top field followed by one bottom field, or one bottom field followed by one top field) and together constitute a coded frame. The two field VOPs that comprise a coded frame shall be encoded in the bitstream in the order in which they shall occur at the output of the decoding process.

When the first VOP of the coded frame is a P-field VOP, then the second VOP of the coded frame shall also be a P-field VOP. Similarly when the first VOP of the coded frame is a B-field VOP the second VOP of the coded frame shall also be a B-field VOP.

When the first VOP of the coded frame is a I-field VOP, then the second VOP of the frame shall be either an I-field VOP or a P-field VOP. If the second VOP is a P-field VOP, then certain restrictions apply (see 7.16.7.4.5).

6.1.1.4.2 Frame VOPs

When coding interlaced sequences using frame VOPs, the two fields of the frame shall be interleaved with one another and then the entire frame is coded as a single frame-VOP.

"

18) Replace the following text in subclause 6.1.3.5,

"

1) the modulo part (i.e. the full second units) of the time base for the next VOP after the GOV header in display order

"

with

"

1) the modulo part (i.e. the full second units) of the time base for the next VOP after the GOV header in display order. For Studio Profiles particularly, SMPTE 12M time code information that is not used by the decoding process.

"

19) Replace the following text in subclause 6.1.3.6,

"

6.1.3.6 Format

In this format the Cb and Cr matrices shall be one half the size of the Y-matrix in both horizontal and vertical dimensions. The Y-matrix shall have an even number of lines and samples.

The luminance and chrominance samples are positioned as shown in Figure 6-1.The two variations in the vertical and temporal positioning of the samples for interlaced VOPs are shown in Figure 6-2 and Figure 6-3.

Figure 6-4 shows the vertical and temporal positioning of the samples in a progressive frame.

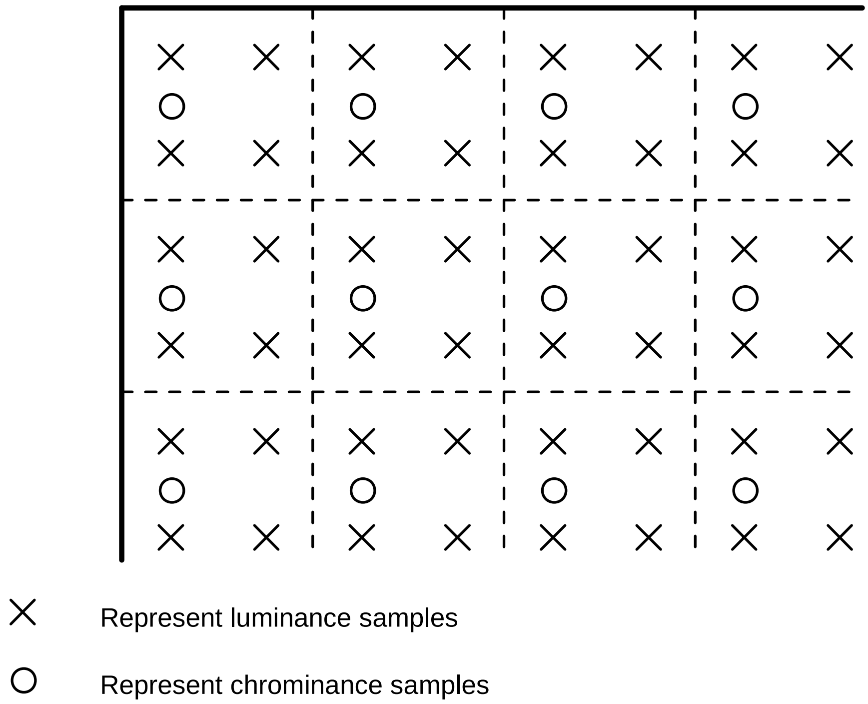

Figure 6-1 — The position of luminance and chrominance samples in 4:2:0 data

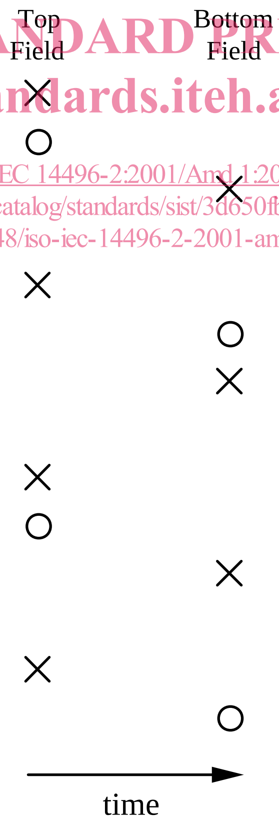

Figure 6-2 — Vertical and temporal positions of samples in an interlaced frame with top_field_first=1

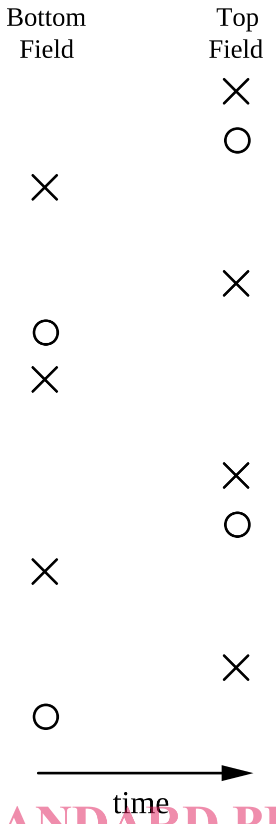

Figure 6-3 — Vertical and temporal position of samples in an interlaced frame with top_field_first=0

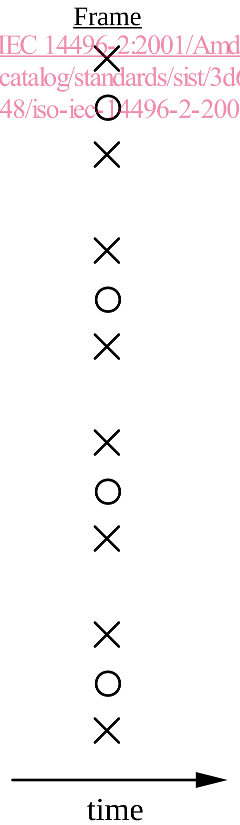

Figure 6-4 — Vertical and temporal positions of samples in a progressive frame

The binary alpha plane for each VOP is represented by means of a bounding rectangle as described in clause F.2, and it has always the same number of lines and pixels per line as the luminance plane of the VOP bounding rectangle. The positions between the luminance and chrominance pixels of the bounding rectangle are defined in this clause according to the 4:2:0 format. For the progressive case, each 2x2 block of luminance pixels in the bounding rectangle associates to one chrominance pixel. For the interlaced case, each 2x2 block of luminance pixels of the same field in the bounding rectangle associates to one chrominance pixel of that field.

In order to perform the padding process on the two chrominance planes, it is necessary to generate a binary alpha plane which has the same number of lines and pixels per line as the chrominance planes. Therefore, when non-scalable shape coding is used, this binary alpha plane associated with the chrominance planes is created from the binary alpha plane associated with the luminance plane by the subsampling process defined below:

For each 2x2 block of the binary alpha plane associated with the luminance plane of the bounding rectangle (of the same frame for the progressive and of the same field for the interlaced case), the associated pixel value of the binary alpha plane associated with the chrominance planes is set to 255 if any pixel of said 2x2 block of the binary alpha plane associated with the luminance plane equals 255.

"

with

"

6.1.3.6 Format

6.1.3.6.1 4:2:0 Format

In this format the Cb and Cr matrices shall be one half the size of the Y-matrix in both horizontal and vertical dimensions. The Y-matrix shall have an even number of lines and samples.

If the matrices represent RGB colour primary matrices, this 4:2:0 format shall not be applied.

NOTE — When interlaced frames are coded as rectangular field VOPs , the VOP reconstructed from each of these field VOPs shall have a Y-matrix with half the number of lines of the corresponding frame. Thus the total number of lines in the Y-matrix of an entire frame shall be divisible by four.

The luminance and chrominance samples are positioned as shown in Figure 6-1.The two variations in the vertical and temporal positioning of the samples for interlaced VOPs are shown in Figure 6-2 and Figure 6-3.

Figure 6-4 shows the vertical and temporal positioning of the samples in a progressive frame.

In each field of an interlaced frame, the chrominance samples do not lie (vertically) mid way between the luminance samples of the field. This is so that the spatial location of the chrominance samples in the frame is the same whether the frame is represented as a single frame-VOP or two field-VOPs.