CCITT SG XV
Working Party XV/1
Experts Group on ATM Video Coding

Delayed contribution
Document AVC-441
January 1993

INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO-IEC/JTC1/SC29/WG11
CODED REPRESENTATION OF PICTURE AND AUDIO INFORMATION

ISO-IEC/JTC1/SC29/WG11
MPEG 93/....
Rome, January 1993

Title: **Proposals for and discussion on the syntax to exploit the correlation between similar video signals over a pair of video codecs**
Source: **PTT Research (NL)**
Purpose: **Proposal, discussion**

Introduction

Assume the reception of two related video signals in a receiver as depicted in figure 1. The two independant input channels represent a simulcast situation, where e.g. channel 2 carries an HDTV signal and channel 1 a TV signal of the same program. The two signals are offered simultaneously, serving also the receivers that are equipped with a TV decoder only. The relation is that one signal is a downconverted version of the other or of a window of the other, e.g. TV with 4:3 aspect ratio from HDTV with 16:9 aspect ratio.

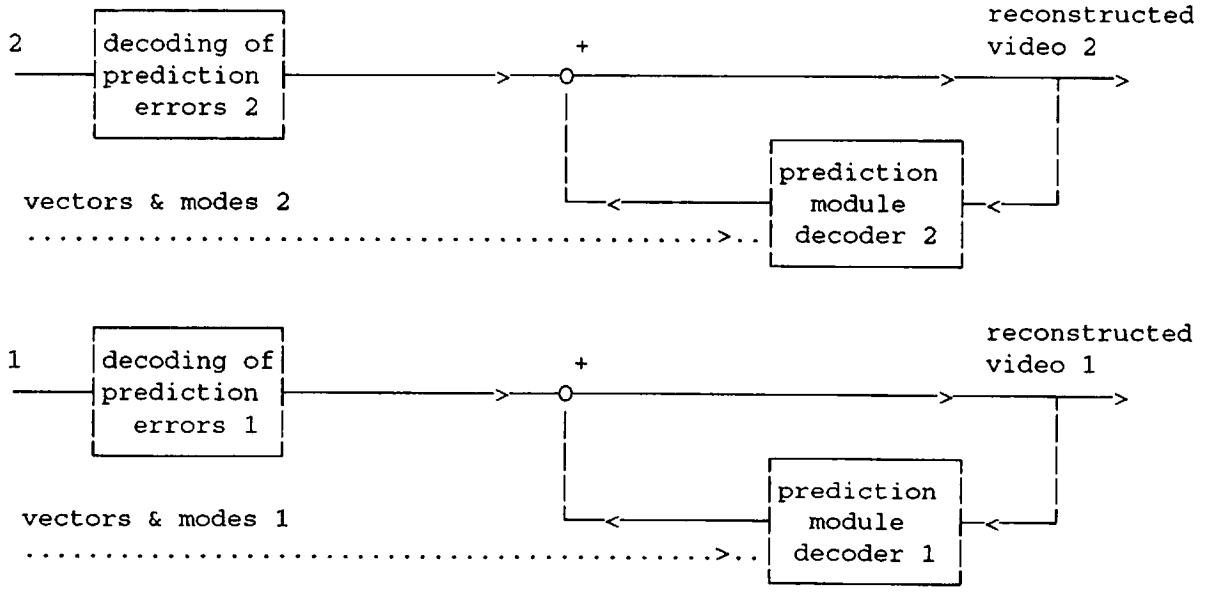

Figure 1: Simulcast situation

Another application is in stereoscopic video (3D-TV), where channel 1 would carry a left view signal and channel 2 a right view signal. Transmission of stereoscopic video signals by the use of MPEG video codecs is currently one of the topics in the European RACE DISTIMA program.

In both applications the video signals 1 and 2 are highly correlated and substantial bandwidth savings can be obtained by exploiting this correlation, especially if the bitrates in channels 1 and 2 are in a similar range, e.g. 2 + 4 Mbit/s.

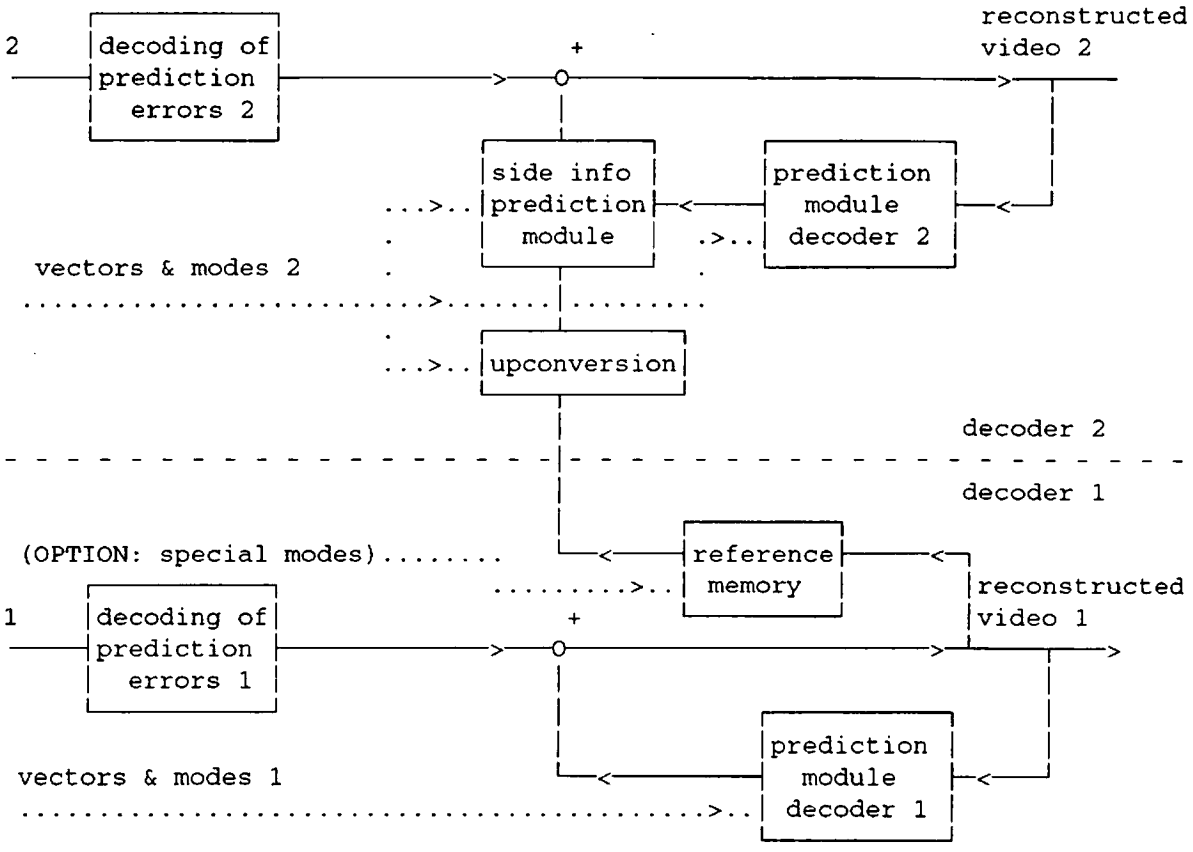

Figure 2: Embedded bitstream situation

In figure 2 the general situation is depicted for embedded bitstream decoding, where decoded pictures of channel 1 serve as a prediction for channel 2. Note that this prediction is retrieved from the reference memory, which typically contains a reconstruction of the downconverted pictures as carried in channel 1. Note also that the pictures may also be of equal size, but have e.g. a lower SNR, or even a progressive scanned signal with an embedded interlaced signal.

Optionally, special modes may exist to fill the reference memory in a way to better serve as a prediction. This would be the case in stereoscopic video coding, to compensate for the different viewpoints of the original left- and right-view vectors. The required additional information could then be transmitted as **user_data** in the picture layer, thus not affecting the general MPEG2 syntax.

We evaluated the capabilities of the current TM3 description to support general embedded bitstream coding as in figure 2 and thereby came to following comments and proposals for modifications and additions.

The channel 2 decoder needs to know which picture from channel 1 is to be used as reference for prediction and how the upconversion should be done, but what type of decoding process is used in the channel 1 decoder is of no interest in decoder 2.

Synchronization of the decoders

Synchronization of multiple video- and audio signals is supported by the MPEG system multiplex. However, the decoding process in decoder 2 of figure 2 needs to be lagging behind the decoding process in decoder 1: picture 1 should be decoded by the time decoder 2 will address it for prediction. When the temporal difference between picture 2 and its channel 1 reference picture is limited to one picture period, we don't think additional temporal reference for embedded decoding of channel 1 + 2 is needed. In any case, one reference memory as shown in figure 2 should be sufficient.

Sequence Header

Note that MPEG2 may support both progressive and interlaced pictures, but this is not indicated in the sequence header. This indication was initially not essential for decoding, but can be very helpful for switchable displays, post processing etc. We are not even sure if in the currently proposed prediction modes this indication can be missed.
We therefore propose:

PROPOSAL 1: include an **interlaced** code just after the **extension_start_code**, indicating whether the source signal was interlaced or progressive. Its syntax is:

interlaced 2 uimsbf,
11 = not used
10 = progressive
01 = interlaced
00 = not specified (default, as in MPEG1 without extension)

The sequence header currently contains the 1 bit **sscalable** code in the extension field. The purpose of this code is to indicate the need of an embedded bitstream decoder (which does not necessarily needs to decode downscaled pictures. For the more general case we therefore propose to rename this bit into **embedded** code, with exactly the same functionality and definition as for **sscalable.**

Further in the current sequence header, in the *if (sscalable) { .....}* section which is to be renamed into *if (embedded) { .....}*, we consider most of the **sscale_code** as superfluous. Decoder 2 does not need to know which coding method, picture format etc. is applied in channel 1, but the system definitely should know which elementary stream should be fed into decoder 1:

PROPOSAL 2: We therefore suggest to start this section with the 8 bit **stream_id** (uimsbf) of the system layer which identifies this elementary stream.

The **sscale_code** will then be replaced by a separate 3 (uimsbf) **ssr** code, with the subsampling ratios as in the current TM3, but with code 0 meaning: no subsampling. NOTE: the **ssr** should be specified referring to the picture sizes of frame pictures, as is the case for the **horizontal_size** and the **vertical_size**, to avoid possible confusion when both field and frame pictures are used.

The other fields in the *if (embedded) {...}* section remain unchanged, specifying the side info prediction module as shown in figure 2.

Picture layer

We think the current syntax elements are appropriate, but we provide following comments for clarification:

COMMENTS: The **lower_picture_reference** code is essential for decoder 2 to know which reconstructed picture of decoder 1 serves as prediction. Also for the **lower_picture_reference** code the unit is frame (input picture). It may also be usefull for synchronization of the two decoders, as one has to be lagging behind the other.

Further: the **overlap_horizontal_left_upper_offset** and **overlap_vertical_right_upper_offset** indicate the picture segment for which the side prediction is available. They are needed to translate the MB adresses correctly.

Macroblock layer

Note that the embedded (compatible) prediction can only be accessed per macroblock. In contrast, for temporal prediction a distinction is made for frame and field pictures, where different motion prediction modes exist. Also a choice between frame based and field based DCT is available. Which modes to use for embedded prediction is still for further study.

CONCLUSION

In the sequence header we propose to ADD a 2 bits **interlaced** code, after the extension code, to indicate the scanning mode to the receiver.

In the picture header we propose to:

1. RENAME **sscalable** in the picture header into **embedded**
2. ADD a **stream_id** codeword in the *if (embedded) {..}* section
3. REPLACE the **sscale_code** by a 3 bit **ssr** code
4. ADD a note that the **ssr** is referring to frame pictures

Further, the general concept of embedded decoding and several syntax elements were discussed.