CCITT SGXV
Working Party XV/1
Experts Group for ATM Video Coding

Document AVC-295
July 1992

SOURCE: Australia

TITLE: Issues concerning the support of Flexible Spatial Resolutions

PURPOSE: Proposal

Abstract

This document builds on previous proposals concerning the necessity for codecs compliant with Rec. H.26X to be capable of carrying a range of picture formats. The need to support a range of spatial resolutions[1], which lie within some maximum resolution limits, was discussed in AVC-219, and the present document discusses the implications of this.

1. Introduction

The Experts Group is developing a video coding system for B-ISDN which should be useful in a wide range of applications. A large number of potential applications have already been identified (see AVC-109 and Rec. I.211) and it is emerging that one of the most significant early applications for video on B-ISDN will be in window-based multimedia information presentation [IVS Baseline Document, June 1992]. In a codec intended for a range of applications, flexibility in certain picture format parameters is essential. It was proposed in Doc. AVC-219 that limited flexibility to support different spatial resolutions should be provided, and that this does not have a significant impact on codec complexity.

While there seems to be agreement that there is a need to support a range of picture sizes, the discussion at the March 1992 meeting concerned the alternative methods of achieving this capability. There appear to be three basic approaches:

- Conversion. All input source picture formats are to be resampled before encoding, so that the encoded picture is always the same size. Recent studies have shown that such intraframe resampling operations can be performed with little impact on perceived picture quality, but that there is a significant penalty in transmitted rate that results from an expansion of the number of pixels to be encoded [AVC-243].
- Flexible support. The input source format is encoded without conversion, and side information is provided to the decoder to allow reconstruction at the original picture size. Conversion for display is a decoder or user option. This approach is proposed in AVC-219.
- Padding. A standard full-size frame in the particular class[2] is padded out to fill the array. More pixels than necessary are encoded, but they should be encoded efficiently because there is no real information in the padded areas. Side information about the original array size is still required if the decoder is to be able to present the original picture size and shape (i.e. appropriately cropped from the transmitted picture format).

Note that the "Conversion" approach is not a complete solution, since it does not allow pictures of different aspect ratio to be supported. One of the other methods will also be required, and side information about the original picture format must still be transmitted.

---

[1] In this document spatial resolution is used to mean the number of horizontal and vertical pixels.

[2] The service classes which are currently being considered are defined in AVC-106R, Annex 3.

2. Flexible Support

To briefly summarise this approach from the description in AVC-219, a set of maximum resolutions should be defined for each service class. These resolutions represent the capabilities of the *decoder*. Coders can then use any resolution, from an appropriately defined subset, within this maximum. The resolution which is used by the coder should be signalled to the decoder in the bit-stream. Figure AVC-295/1 illustrates the concept.

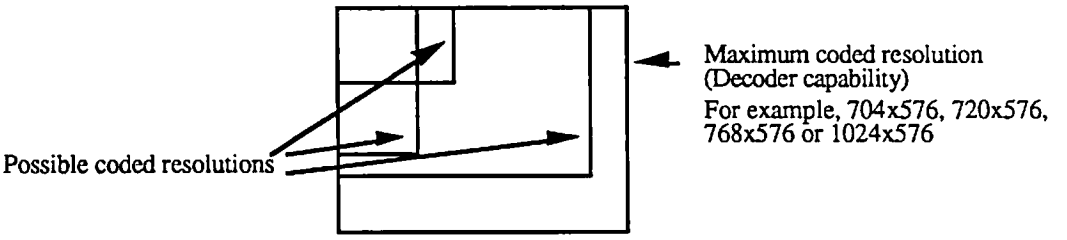

Figure AVC-295/1. Illustration of the flexible coder formats proposal.

3. Padded Frame

This approach assumes the restriction of only one possible picture array size that can be transmitted for a particular class. Therefore, any picture that is smaller than this size must be padded out to fill the array, as shown in Figure AVC-295/2. Some unnecessary distortion may be introduced at the picture boundaries because of the artificial edges introduced, but this could be avoided if the smaller picture contains an integral number of transform blocks (or if pixel value extension is used, but this would also contribute to the transmitted rate). Encoding and decoding of a larger array size than necessary will mean that more processing is required at both ends of the link, and this could have implications when the processor could be shared with other tasks or where a dedicated application is involved.

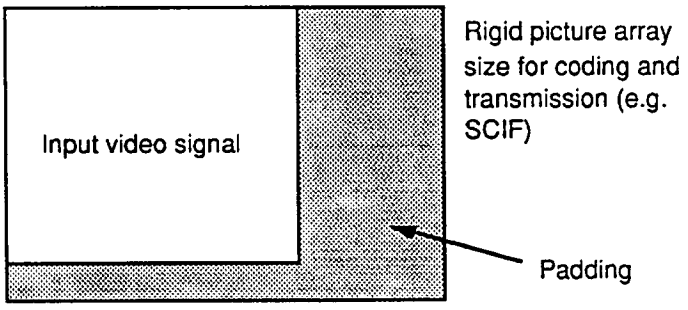

Figure AVC-295/2. Illustration of the padded frame approach.

4. Comparison and Impact

4.1 Side Information.

As noted in Section 1, both the flexible picture format and padded frame approaches require the (reliable) transmission of the original image parameters. In the flexible format approach, these parameters will be used to determine stripe/slice sizes, etc. as well as defining a window size for presentation on screen. In the padded frame approach, it is just the window dimensioning that requires this information.

4.2 Encoder

In the padded frame approach, the encoder always codes the maximum number of pixels for that class, although it may be used only with a single lower resolution source. While a decoder may be used to access information from a range of sources (and this is the basis of the unified approach sought by the SGXV EG), it is much more likely that an encoder will be dedicated to a single source

such as a camera. If the required image size is smaller than the full padded array (an example would be a videophone application with a 1:1 aspect ratio), then there is an unnecessary processing burden since dummy data will have to be checked to see if there is ever any useful information and the total pixel throughput rate is higher than necessary. An alternative would be to customise the encoder for each application, so that it knows which blocks of the picture array must always be labelled as "not coded". Such an approach would seem to be contrary to the aim of having a coding device that could be applied widely. The processing burden may not matter if single mass-produced encoder chips are used but, if multiple general-purpose processors or (as will become increasingly important) software solutions are used, then a cost or performance penalty may result. The rate overhead is not likely to be large for most frames, since uncoded blocks may just be skipped and the penalty is only seen in the codes used for differential block addressing. However, in MPEG I-frames, there is no such thing as an uncoded block, and an overhead of at least 30 bits/block would result (about 24 kbits overhead for each I frame for a half-padded Rec. 601 picture).

The flexible format approach processes no more pixels than necessary, so dedicated encoder solutions could use the minimum number of processors or could free a shared processor to perform other operations. There is also no overhead in transmitted rate.

4.3 Decoder

The same issue of processing burden arises when considering the decoder, except the decoder could be used to process compressed signals coming from a variety of sources and with a variety of characteristics. While a decoder is less likely to be dedicated to a particular application, it is perhaps more likely to be using processing power that is shared between multiple operations. For example, the one processing module may be used to decode several simultaneous video bit streams for continuous presence multipoint videoconferencing. Software decoding will become increasingly important on desktop PCs or workstations.

The flexible format approach will use minimum processing resources, freeing them for use in other applications. The padded frame approach has only a small processing overhead, but must generate more pixels than are used (and which later may be cropped for presentation).

Again, if dedicated single chip decoders are used, there is no penalty in complexity for either system.

4.4 Impact on Picture structure

If a single picture spatial resolution is used for all video transmission (within a class), the system could work with a rigid block grouping structure such as the GOB in H.261. A flexible-width stripe structure, such as that used in MPEG, is required for the flexible format approach. This appears to offer no significant penalties, since the picture dimension parameters (for use in indexing and motion estimation/compensation) will simply be read from registers instead of being always fixed. It has already been noted that the picture parameters must be transmitted reliably in either case.

Since the current aim of the SGXV EG is to achieve common text with MPEG for the coding aspects of H.26X, it is important to consider the MPEG picture structure. MPEG2, intended as a generic coding scheme, can be expected to retain the flexibility of picture dimensions that is a feature of MPEG1. The capability of the underlying coding/decoding algorithm to process varied resolution pictures is therefore almost certain.

5. Conclusion

The requirement to process pictures with parameters up to some maximum for the class is necessary to provide for interworking between many applications and terminals.

While studies on conversion techniques to a common format for coding and transmission are continuing, this cannot provide a total solution and there will remain the problem of transporting picture formats which do not completely fill the class sizes. Some means of accepting these signals

and ultimately displaying them in windows of the original dimensions will be an essential feature of H.26X if it is to satisfy the goal of being applicable to a range of applications.

The above discussion has indicated that coding of the original picture dimensions (flexible picture formats) has the following features:

- No conversion is required, so the associated delay and complexity penalties are avoided;
- Encoders dedicated to particular applications (i.e. particular input picture formats) can be implemented with minimum processing load and without the need for customisation to maintain maximum coding efficiency;
- Decoders can operate with the minimum processing load for the particular incoming signal, freeing resources for other tasks or other signals;
- While the transmitted rate overhead and compression processing increase for the padded frame approach are likely to be relatively small (except for I-frames), this solution would require the input processing and output generation of pixels at a greater rate than necessary. These can all be avoided with the flexible approach;
- The required picture format flexibility (flexible sized stripes) is consistent with MPEG.

In view of the above considerations, Australia proposes that Rec. H.26X should directly accept, encode and transmit video signals with a range of spatial resolutions. Restrictions on the allowable range (due to, for example, the chrominance subsampling structure or sizes that are multiples of the block size) are for further study and discussion.