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

Document AVC-36
May 1991

SOURCE: Australia

TITLE: Picture Formats Supported by Video Codecs on the B-ISDN

PURPOSE: Proposal

Abstract

This document considers the picture formats that should be used in the coding of video services for the B-ISDN. An early decision on the picture formats which a coder will be capable of dealing with is essential if integration of video services, by whatever means, is to be achieved. A coder must support existing picture formats and allow interworking between video services. The contribution concludes by proposing codec capabilities which satisfy these requirements.

1. Introduction

The ATM experts group is aiming to define a unified coding algorithm which can be used for applications in a broad range of service classes. It has been agreed that maximum integration of video service delivery should be facilitated by the coding algorithm. This implies interworking across the service classes. Although the means by which interworking will be achieved has not been decided, consideration of a number of alternatives [AVC-35] suggests that, no matter which is chosen, it is necessary to define a family of codec picture format capabilities which is compatible with a range of source formats in all the service classes.

The family should be defined in such a way that it:

- Maintains qualities at least equal to those which presently exist in defined services;
- Allows flexible interworking between service classes;
- Facilitates video coding using state of the art techniques.

The flexibility of the B-ISDN is likely to lead to a rapid expansion in the number and type of video services. For example, multi-media communication will involve coding of images and video sequences at a broad range of resolutions and frame rates which may not precisely coincide with any of the existing services. Hence flexibility in the definition of service resolutions is essential.

In the following section a discussion of the parameters which must be chosen and the issues which impact on appropriate choices is presented. A family of video signal parameters is then detailed which satisfies the goals of maintaining quality, allowing interworking and facilitating coding.

2. Discussion: Parameters defining Video Signals.

2.1 Pixel Shape

Within existing recommendations a single pixel shape has yet to be standardised. The defined formats for MPEG and CCIR Rec. 601 allow two possible shapes which coincide with the two main television standards (NTSC, PAL/SECAM). With the likely integration of video communications and desktop workstations it seems desirable to seek some compliance with the computer industry and use square pixels. Future digital HDTV video standards may also use square pixels. A further argument for reconciling the pixel shape definitions is that certain camera and display technologies (e.g CCD cameras and LCD displays) have their pixel shape fixed at manufacture. A single shape for all applications would result in reduced production costs due to economies of scale.

Multiple standards limit interworking capabilities. Problems will certainly arise with international communications involving different standards, and conversion will be required at some point. The question really is where this conversion should take place and whether it should be a part of a defined standard. The options are

- Define a single standard pixel shape and therefore standardise conversion methods from existing formats.
- Permit multiple pixel shapes and allow the conversion processes required for interworking between different standards to be developed as add-on capabilities.

The first option is restrictive and would necessitate the inclusion of more complex conversion algorithms in the defined coding architecture. Since the precise pixel shape does not affect the coding structure choosing the second option does not increase coder complexity. The pixel shape does influence the number of pixels horizontally and vertically for a given aspect ratio. The spatial resolutions must therefore be chosen to accommodate all permissible pixel shapes at the required quality levels.

2.2 Spatial Resolution

In defining a set of resolutions we need to make sure we can:

- Facilitate coding.
- Cope with as many existing formats as possible.
- Allow flexible interworking between a range of services.

As suggested in the introduction, maintaining flexibility in the definition of resolutions is important so that a wide range of video sources can be dealt with. In most currently defined coding techniques the 8x8 DCT plays a central role in redundancy removal. Block transform techniques are likely to remain important in future codecs and to facilitate their use it is important that defined resolutions contain an integer number of transform blocks. The fact that the colour representation is likely to be sub-sampled with respect to the luminance, in at least one dimension, by a factor of two makes it desirable that the sets of allowed luminance resolutions be multiples of sixteen (however other sets based, for example on multiples of 8 would be possible). Within this constraint it is possible to construct codecs which can cope with a wide range of resolutions and aspect ratios.

For particular applications, there is no single spatial resolution which is universally employed. For example the CIF video-conference format can be 352x288 or 352x240, depending on whether the originating source is based on PAL/SECAM or NTSC. As a result we must be careful to ensure that, in defining coders for particular applications, the coders are capable of coping with all spatial resolutions which may be used for the application. To permit flexibility, coder resolution limits are proposed for each service category. Any service in a particular category will have a resolution equal to or below this figure.

The following set of coder resolution limits is proposed (luminance specified):

1.	Video-phone:	≤192x144	(includes QCIF, 176x144/120)
2.	Video-conference:	≤384x288	(includes CIF, 352x288/240)
3.	TV:	≤768x576	(includes CCIR Rec. 601, 720x576/480)
4.	HDTV:	≤2048x1152	(includes all HDTV proposals, 1920x1152/1080/1035)

These resolutions are large enough to include all defined formats for each service. They also permit square pixel shapes at the appropriate aspect ratio (4:3 for services 1-3 and 16:9 at service 4). A coder capable of operating at these resolutions is able to accommodate all existing video formats, so manufacturers would be encouraged to produce equipment which meets these specifications. This also ensures that all coders developed for a particular application will be able to interwork, regardless of the input source resolution. Note that the above list of services is not intended to be exhaustive. Other services, such as extended definition TV (960x576/480), may be important video services on the future B-ISDN. The general principle of defining coder resolution capabilities which encompass all defined formats for a particular service is important for maximum integration and should be applied to all services.

Conversion between the resolution levels is required for integration across services and for the definition of layered coding systems. To simplify service integration it is useful to have coder resolutions related by integer or rational factors. For coder resolutions 2 and 3 a lower layer resolution is obtained by decimating by two in each direction. For the HDTV coder resolution an allowable TV coder resolution can be obtained by decimating by 8/3 to provide a letter box format (all picture information displayed in a window). If side panel information is discarded it is possible to use a decimation by two.

An example may help to illustrate the way the system would work. A video-conferencing terminal capable of decoding 384x288 pixel images could accept any resolution that is a multiple of sixteen in each direction, so the existing 625-line based CIF image format (352x288) or the 525-line version (352x240) could both be supported. Furthermore, the same terminal could also cope with computer generated images of 256x256 pixels, or any other resolution included in a 384x288 pixel array, which is a multiple of sixteen.

2.3 Luminance/Chrominance Sampling Pattern

It is preferable to be able to define a single sampling pattern which can be used for all services. Again the problem is that we have a number of existing services, with different sampling patterns, which the unified coding scheme should encompass. The two main formats which need to be dealt with are: CIF/QCIF used in Rec. H.261 and MPEG1; and the CCIR Rec. 601. CIF/QCIF uses a pair of chrominance samples for every four luminance samples. CCIR Rec. 601 uses a pair of chrominance samples for every two luminance samples. Both have orthogonal sampling lattices.

The problem is to define a format for enhanced quality services which encompasses CIF/QCIF and CCIR Rec. 601. A full interworking codec which can cope with multiple formats is likely to be quite complex. Different ways of dealing with motion compensation of the chrominance blocks would be required as well as output conversions for the display device. A single standard may not be well suited to all applications (e.g. both contribution and distribution of TV). This aspect of the video signal format is likely to be one of the more difficult to resolve. It is important that the experts group seek input from CMTT, ISO/IEC and CCIR SG11 before reaching a final decision.

2.4 Scanning method

One of the most difficult problems to deal with in defining video formats is the fact that the input source scan format is interlaced (most cameras and all recorded material is interlaced), while coding is simplified if the format is progressive. This is not a serious problem for the Rec. H.261 and MPEG1 coders since it is only necessary to process the odd fields in a CCIR Rec. 601 video signal. Strict quality requirements which may exist in certain applications on the B-ISDN require some care in defining the scan format and/or the conversion methods.

To avoid undue coder complexity a single scan format should be chosen. Maintaining compatibility with existing video codec picture formats suggests that a progressive scan should be proposed. The main difficulty with this choice is that, although most predictions are that progressive scanning is universally preferred in the long term, TV and, in the near term, HDTV formats will be interlaced. A conversion between interlace and progressive scan formats would be required if a progressive format is adopted for coding. The conversion from interlace to progressive and back will have implications on signal quality and may make it difficult to satisfy the requirements of broadcasters, particularly for contribution quality coding. CMTT have suggested [CMTT-2/TEMP/7-E, AVC-28] that a unified coding scheme should be able to deal with both interlace and progressive signals. Again it is essential that the study group seek input from CMTT, ISO/IEC and CCIR SG11 on a suitable definition for the scanning format.

2.5 Temporal Resolution

Different frame/field rate standards continue to exist in different parts of the world. Interworking requires that families of frame rates be defined which are integer multiples of some basic rate. For example, 12.5, 25, 50 Hz or 15, 30, 60 Hz. Interworking between families can be assured by developing flexible coders and displays which will cope with a range of different frame rates. This is possible with little additional cost/complexity and should be encouraged.

In developing coders for particular applications it is important to define coder frame rate capabilities which meet the quality expectations for the application. These should encompass all the existing standards for a particular service in the same way the spatial resolution figures encompass existing formats. The following set of coder frame rate capabilities is proposed:

1. Video-phone: ≤15 Hz (encompasses 12.5 Hz European/Australian and 15 Hz USA/Japan)

2. Video-conference: ≤30 Hz (encompasses 25 Hz European/Australian and 30 Hz USA/Japan)

3. TV: ≤30 (possibly 60) Hz (encompasses 25/50 Hz European/Australian and 29.97/59.94 Hz USA/Japan)

4. HDTV: ≤60 Hz (encompasses 50 Hz European and 60 Hz Japan)

The frame rates for video-phone and video-conference services encompass the current QCIF and CIF standards. It is unclear at this stage what an appropriate frame rate for a TV service is since it may depend on the choice of scan format. Further input is required from CMTT, ISO/IEC and CCIR SG11.

2.6 Colorimetry:

For most existing and near term applications the colour primaries used in the two main television systems should be adequate. However there is a possibility that future high definition services will utilise colour primaries providing a wider colour gamut. The precise definition of the primary colours does not directly affect the coding architecture. Conversion between colour primary systems is a straightforward matter and therefore the existence of source material using alternative definitions should not present problems. In the long term, a set of primaries offering a wider colour gamut would be attractive for a coder which is to be capable of delivering a broad range of services. CCIR SG11 is active in this area and their input is required.

3. Proposed Codec Format Capabilities

Based on the arguments raised in the previous sections the following parameter definitions are offered for further discussion.

Coder resolution capabilities:

Video-phone:	≤192x144 luminance, ≤15Hz frame rate
Video-conference:	≤384x288 luminance, ≤30 Hz frame rate
TV:	≤768x576 luminance, ≤30 Hz frame rate
HDTV:	≤2048x1152 luminance, ≤60 Hz frame rate

Within each service class luminance resolutions which are multiples of sixteen should be allowed, however for some service classes a more flexible definition based on multiples of eight may be desirable.

Pixel Shape:

1:1 (though services with a variety of pixel shapes, including the present PAL/SECAM and NTSC definitions, can be supported)

Sampling relationship:

Orthogonal sampling. Further input from CMTT, ISO/IEC and CCIR SG11 should be sought before definition of the Luminance/Chrominance sampling relationship.

Scanning:

Further input from CMTT, ISO/IEC and CCIR SG11 should be sought.

Colorimetry:

Luminance and two colour difference signals using existing primary definitions. Enhanced colour representations are for further study and would benefit from inputs from CMTT, ISO/IEC and CCIR SG11.

4. Conclusion

The determination of a set of video signal parameters that coders operating on the B-ISDN should be capable of processing is a necessary first step in defining a coding architecture which can be applied to a broad range of services and provide flexible interworking. Therefore it is imperative that the Experts Group reach an early decision on a suitable set. The issues raised in this contribution should be taken into account in reaching this decision and the parameters suggested can be used as a basis for further development. Note that the figures proposed in this document are limiting capabilities and are not intended as redefinitions of any existing service parameters. The formats should be agreed with other relevant standards bodies to facilitate integration of video services on the B-ISDN. Liaison statements and a contribution to the IVS Baseline Document are recommended.