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

Document AVC-73
August 1991

SOURCE: Australia

TITLE: Architecture for Video Service Provision

PURPOSE: Discussion

Abstract

A general architecture for B-ISDN video service provision is outlined. This architecture provides sufficient flexibility for a range of compatible, interworking codecs. It is suggested that such an architecture is the only means of providing for compatibility with existing coding schemes in the short term, while allowing for coding schemes optimised for the B-ISDN. The architecture is layered in the sense that several separate video coder format resolution limits are identified. The importance of reaching a decision on the coder input formats defining layer boundaries is highlighted.

1. Introduction

It is clear that, in defining a video coding architecture for the B-ISDN, various forms of *compatibility* are key requirements. *Forward*[1] and *backward* compatibility are important for the evolutionary introduction of new services. The Experts Group have agreed that maintaining compatibility between any new coding system and existing systems should be given high priority [AVC-28]. The CCITT Rec. H.261 codec is seen as particularly important since interworking between H.261 codecs on the narrowband ISDN, and any new video coding system for the B-ISDN, should be possible. *Upward* and *downward* compatibility are required for interworking between coders designed for different services on the B-ISDN [1, 2, AVC-35]. Task Group CMTT/2 [1] have stressed that it is important to have compatibility between conventional TV and HDTV digital distribution from the beginning. They have also stressed the importance of being able to extend HDTV in a compatible manner [1]. The same argument can be put forward for communicative services, where interworking between low and high quality video-phone/video-conference terminals on the B-ISDN will be important for wide acceptance of the services.

Maintaining forward/backward compatibility is always a short term goal and providing these compatibilities should not compromise the ability of the coding architecture to evolve as the compatibility requirements change. The requirements for upward and downward compatibility will also change as the resolutions at which different services operate change. Providing a coding architecture which satisfies the current compatibility and interworking requirements, while at the same time allowing future enhancements, requires a great deal of flexibility. In this document a suitable architecture is proposed.

2. Architecture

The flexible layering architecture for service interworking [AVC-35] has been recognised as a useful idea which has a number of advantages in terms of compatibility and interworking [AVC-65R]. This architecture allows flexibility in the number and arrangement of layers which are employed in a layered coding system. It includes both single layer and fully layered coders as special cases [AVC-35]. However, this architecture, on its own, cannot provide the compatibility with existing systems which may be required. To cater for interworking with a number of defined coding algorithms the concept of flexible layering must be extended to allow flexibility in the coding techniques which are used for the layer signals. This extension to the flexible layering architecture is illustrated in figure AVC-73/1.

[1] Forward, Backward, Upward and Downward compatibility are defined in AVC-32.

The vertical axis in figure AVC-73/1 represents video signal resolution (spatial and temporal). This axis is divided into a number of separate regions which define the layers. At this time it seems clear that the lower layers in this structure will code CIF/SIF/QCIF format signals, however the format of the upper layers is still under discussion. The flexible layering structure is quite general and can be used to develop a coding architecture no matter which high quality formats are chosen, however, the choice is likely to impact both coder complexity and efficiency.

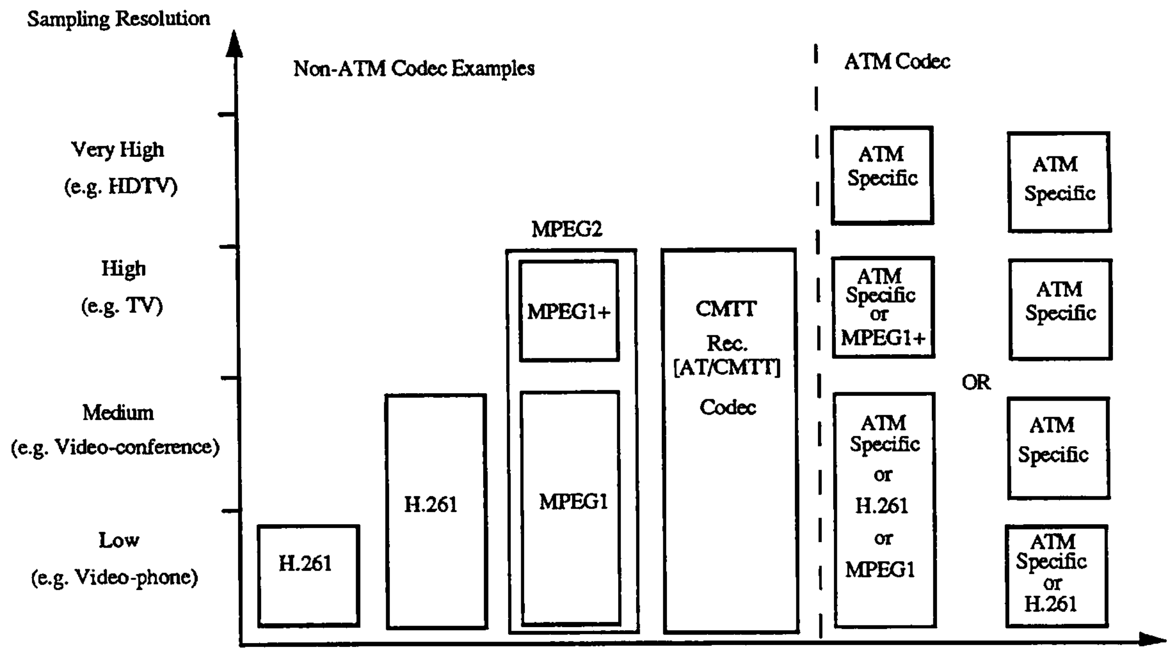

Figure AVC-73/1. Flexible architecture for video service provision.

In the diagram, coders which extend from a particular layer down to the base operate on full-band video signals. Coders which span only middle layers code incremental signals. There are a number of well defined coders which could be used to code full-band video signals. Several examples of existing coding techniques which could be mapped onto this structure are shown. These include single layer coders such as an H.261 coding at the QCIF or CIF level and an MPEG1 coder at SIF resolution. The MPEG2 algorithm is yet to be determined. It could be designed so that forward/backward compatibility with the MPEG1 scheme is maintained (by using a coder labelled MPEG1+ to code the incremental layer), or as a non-compatible single layer system.

H.261 and MPEG algorithms have not been designed for the B-ISDN and it is likely that an ATM specific coder will be required for optimum coding, as shown. However, bit-stream compatibility with the existing schemes may be important. From previous SGXV documents [AVC-33, AVC-58], the possibility of switching between H.261, MPEG1 and a new ATM specific coding algorithm to provide bit-stream compatibility is realistic. The basic building blocks used in both the H.261 and the MPEG coders have much in common, since the techniques used for coding in MPEG include, as a subset, the techniques used in H.261. It seems likely that the coding techniques at the heart of the new ATM codec will be similar, though some enhancements to cater for the unique characteristics of an ATM network (for example multiple priorities, VBR channels, cell loss), will be required. Hence, it should be possible to provide switchable compatibility with H.261, MPEG1 and possibly MPEG2. This does not imply that all codecs constructed for the B-ISDN will have switchable compatibility with all existing coding schemes, however the feature can be provided, if it is necessary in particular applications.

Of course, interworking between an ATM specific coder on the B-ISDN and an H.261 codec on the N-ISDN could also be provided by performing some sort of translation[1] at the bridge-point between the networks. The preferred approach will depend on:

- The relative cost of providing translation capability in the bridge or switchable coding capability in the codecs.
- The performance/quality achievable if the bridge has to perform significant transcoding (e.g. delay).
- The performance/quality achievable by using a non ATM based coding method over the ATM network (the switchable coder approach could see an H.261 bit-stream transported in cells on the ATM network).

One of the most important properties of flexible layering is that lower layers can be switched off as required. Using the ATM codec shown on the right in figure AVC-73/1, which implements flexible layering, it should be possible to obtain compatibility with H.261 at both medium resolution and at a low resolution. This involves a selection between coding of a medium resolution signal as a single layer, or as two components (low resolution plus enhancement). The ability to selectively decompose layers is necessary to allow video services to migrate to higher resolutions [AVC-35].

An implementation of flexible layering, using a pyramid structure, is given in Appendix 1.

3. Coder Input Formats

To make progress on an overall architecture for video service provision it is necessary to define the coder input formats. This is one of the more difficult issues facing the Experts Group. Below some of the issues which impact on the decision and a number of important requirements which must be satisfied by any choice are identified.

3.1 Spatial Resolutions and Frame Rates

As suggested in [AVC-36], it is essential that the precise resolution at which coding takes place should be flexible. The concept of a flexible resolution has been used in MPEG1 and may be a part of the final MPEG2 coding scheme. Flexibility is important if the coder is to be suitable for a range of source formats. In a broadcast TV application the codec should be capable of standard TV at 4:3 aspect ratio and both Enhanced and High Definition TV at a 16:9 aspect ratio. It should also cater for the regional variations in line number/field rate that exist within these general service classes. Film comes in a range of aspect ratios from 1.33:1 to 2.33:1 and the quality of a broadcast or interactive service would be enhanced if these formats could be dealt with directly. Desktop video communications will require the coding and display of video in a window on a workstation which could be of any size. Multi-media communications will almost certainly require the coding of computer generated animation which could be of arbitrary aspect ratio and resolution. All these different applications, and any new applications which are sure to arise as video communications becomes more widespread, can only be effectively provided if the coded resolution is flexible.

This flexibility should be extended to frame rates. It is within the scope of present technology to implement displays which can deal with a range of frame rates. The source formats used in different regions throughout the world differ in their frame rate and the coding system should be able to deal with all possibilities. For example, interworking between a system operating at 25 frames/sec and one operating at 29.97 frames/sec should be possible with little additional complexity. Both would code at their own source rate and display at the source rate of the incoming material. This would avoid unnecessary frame rate conversion.

---

[1] Here translation involves the necessary modification to the bit-stream to permit communication between non-identical coders. This may or may not involve transcoding (full decoding and coding).

3.2 *Interlace/Progressive Scanning*

It seems clear that, if interworking with MPEG1 and H.261 is important, then the coder format for the lower two layers at least will have to be progressively scanned. What is unclear, at this stage, is whether the coder formats at the higher layers should be interlaced or progressive. The flexible layering architecture which has been proposed could operate using different scanning formats at each layer. There are a number of factors to consider in the scanning format choice, some of which have been raised in previous documents. In summary the most important are:

- All source formats at the present time, and in the near future at least, will be interlaced[2]. If a progressive coding format is chosen a conversion between the interlaced source format and the progressive coder format is required (Progressive displays are common in computing applications, and are appearing in consumer applications, and conversion back to interlaced for display may not be required). Conversion adds complexity to the complete coding system.
- Conversion from interlaced to progressive may lead to unacceptable quality degradations for high quality signals and may not satisfy the requirements of broadcast TV or HDTV.
- Coding of progressive signals is more straight-forward than coding interlaced signals. However, it isn't known whether, for a given quality, coding an interlaced source by converting it to progressive first is more efficient than coding the signal in an interlaced format.

One of the over-riding factors in this decision is the quality of the final service which can be provided. If picture quality cannot be maintained with an interlace to progressive conversion, then an interlaced coding format is required. If, however, acceptable quality for all the required applications can be maintained using a progressively coded format then a progressive coding format would be the most appropriate choice, given the interworking advantages and the desirable long term objective of a unified set of coding formats. Investigation of this issue is important, and views should be sought from CCIR SG11.

3.3 *Pixel Shape*

Another important difference between regional picture formats is the pixel shape. Conversion from one pixel shape to another is relatively straight-forward [AVC-36] and does not introduce any significant picture degradation [AVC-46]. The choice is between a single pixel shape used within the coding system, potentially implying conversion at both the coder and decoder, or multiple shapes with multiple conversion capabilities required at the decoder for inter-regional communications.

If a single pixel shape is chosen, then it should be square. This will encourage the integration of video communications with desktop workstations. A number of display technologies have a pixel shape which is fixed at manufacture (e.g. LCD displays) [AVC-36, AVC-60] and therefore having a common pixel shape for both computer workstations and video communications will lead to overall cost savings. This issue will become increasingly significant as the telecommunications, computing and broadcasting industries continue to converge.

4. Conclusions

Flexible layering has been recognised as a useful concept which provides advantages in terms of compatibility and interworking [AVC-65R]. It has the potential to provide a high degree of backward/forward compatibility with existing coding schemes such as H.261 and MPEG1. To exploit this potential it is necessary to extend the flexible layering concept to allow a choice of

---

[2]Note that in the future, digital television, especially high definition television, signals presented to a codec may have undergone preliminary source coding and processing.

coders at each of the layers. This flexibility also allows for the development of a coding scheme, which will be better suited to video coding on the B-ISDN.

An important first step in developing an overall architecture for video services on the B-ISDN is the definition of a suitable set of coder input formats. This decision seems relatively straight-forward for the low resolution services, where a progressive format is the obvious choice. The choice of format for high resolution services is more difficult. It has been stressed that the precise spatial resolution and frame rate used for coding must be flexible if the coder is to be suitable for a range of material and a wide range of current and future applications.

References

[1] Liaison Statement to CCITT SG XV from TG CMTT/2, 28 March, 1991, in AVC-28.

[2] IVS Baseline document.

Appendix 1. Example Implementation

A number of schemes which can implement the flexible coding architecture which has been described are currently under investigation in Australia. One such architecture, which seems promising, is based on a pyramid decomposition of the input signal. A three layer system is shown in figure AVC-73/A1. This type of interworking scheme has also been described in [AVC-64] and layered architectures, such as this, have been proposed for MPEG2 coders which can interwork with the MPEG1 coding system.

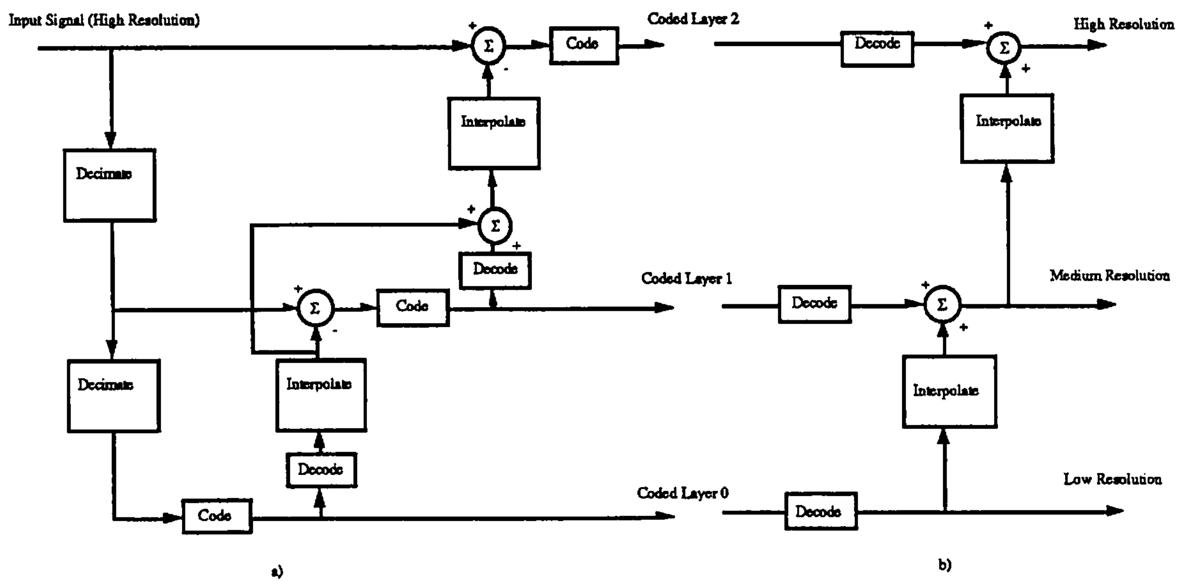

Figure AVC-73/A1. A three layer flexible layering system. a) Coder. b) Decoder.

The architecture is capable of switching from a three layer codec, providing High, Medium and Low resolution, to a two layer codec for providing either High and Medium resolution or High and Low resolution layers. It can also operate as a single layer codec at High resolution. This is achieved by switching off either the layer 0 coder, the layer 1 coder, or both. Hence, if the input to the scheme was a TV quality signal of around CCIR Rec. 601 resolution, then the system could provide a single layer TV quality codec or a two layer TV codec with a CIF or QCIF lower layer or a three layer codec with both CIF and QCIF lower layers.

The coding schemes shown at each layer could be capable of some current standard (e.g H.261 or MPEG1 for the two lower layers if these layers have QCIF and CIF resolution) as well as a technique which is specifically designed for the particular layer signal statistics and suited to the ATM network.

An important question is whether using such a structure impacts on the efficiency of the coding scheme. If compatible coding schemes are used on any of the layers, coding efficiency is likely to be affected. Coders specifically designed for baseband signals will not necessarily code the incremental layer signals effectively. In a companion document [AVC-74], it is shown that, in a two-layer coder which codes CIF and QCIF, the additional lower layer does not, in principle, introduce any additional coded information. This suggests that schemes designed specifically to code the layer signals may be as efficient as single layer coders.