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

Document AVC-35
May 1991

SOURCE: Australia

TITLE: Architecture for Video Service Integration on the B-ISDN

PURPOSE: Proposal

Abstract

The broad integration of video services is an agreed aim that emerged from the first meeting of the Experts Group, and is a key CCITT objective for video service provision on the B-ISDN. This document identifies more precisely what integration means, what the benefits are, how it can be achieved and what decisions are needed to promote it. An architecture is proposed which facilitates service integration without imposing restrictions on particular service categories. A decision on the picture parameters that will facilitate service integration is recognised as essential, independent of the method used.

1. Introduction

The terms of reference of the Experts Group for ATM Video Coding include the study of the "feasibility of a unified coding standard for various applications in all service classes" (AVC-2), and the report of the first meeting (AVC-22R) records agreement "that the Group aims at a universal coding algorithm in terms of services, quality, resolution, applications and bit rates each of which are given a range". Integration of video services is seen as a key objective for their provision on the B-ISDN (see CCITT Rec. I.211 and the IVS Baseline Document AVC-25) and there is also interest in the interworking between MPEG1 and MPEG2 video systems for Digital Storage Media (DSM).

Against this background of broad aims, the work can now be progressed further by considering more specific objectives and the methods by which they can be achieved.

2. Integration of Video Services on the B-ISDN.

Integration of video services is provided by commonality of the manner of their delivery. The B-ISDN provides several components of this integration, since it will deliver all video services over the one transmission medium, using the one transport method (ATM cells), and with the same connection control and signalling systems. The "missing" element is commonality of the video coding system used in different applications.

Video services can be categorised according to:

- the picture parameters used (e.g. QCIF, CIF, CCIR Rec. 601, HDTV signal formats);
- the service parameters (e.g. point-to-point interactive, retrieval, point-to-multipoint, distributive);
- the rate (and associated cost, complexity and quality) at which the coder is operated. For example, the same videoconferencing codec could be operated at around 200 kbit/s or 2000 kbit/s, depending on particular requirements, number of participants, etc.

Early decisions in the evolution of a video coding system for the B-ISDN must include the extent to which these various systems should use an integrated coding approach. Broad integration across the whole spectrum could offer many benefits. Such commonality would provide:

- single terminal access to many services;
- the ability to switch between different quality levels during a connection (for low-resolution browsing, or changing from one to multiple parties displayed in a multipoint conference);

- easy evolution to new services (e.g. HDTV), with a guaranteed user base able to access the programme material;
- easier combination of image/video signals from a wide variety of sources. Currently most common in the TV studio, this will become a desk-top activity with the growth of multimedia services;
- the flexibility to offer continuous presence multipoint videoconferencing through combinations of video signals with differing resolutions, etc.;
- easy upgrading or migration of a particular service class (e.g. videoconferencing at TV quality);
- convergence of computer imaging applications with broadcasting and telecommunications;
- easy mode conversion (for example, from conversational to stored video);
- easier interworking with lower rate networks (e.g. LANs, MANs);
- efficient transmission through selective routing of elements of the video signal through nodes of the network, so that terminals only receive those elements that they intend to decode.

3. Alternative Approaches

There are three basic approaches to provide at least some of the flexibility required for effective integration of video services:

Negotiation Approach:

(This approach is called a "switchable encoder" in CMTT/2-18. See AVC-28).

At the commencement of a connection, terminals negotiate a set of parameters with which both can cope. A set of standards of increasing quality would be defined and a basic capability assumed for all terminals. In multi-party calls, such a procedure would mean that communication is always carried out at the quality of the least capable terminal. This would not be acceptable for broadcast services, would be extremely limiting for multipoint conferencing and would eliminate large portions of the range of video applications that could be integrated. Receiving terminals must have the flexible capability to reconfigure to the various ranges of image resolutions, sampling patterns, frame rates, etc.

Simulcast Approach:

Transmitting terminals contain multiple encoders, operating at a variety of resolutions and quality levels so that broad interconnectivity can be achieved. In a broadcast or point-to-multipoint connection, the different video signal quality levels are transmitted simultaneously. There are several interpretations of "simulcast". For example, each receiver could contain a single decoder, which ensures simplicity but limits flexibility, or or it could contain multiple decoders allowing selection from the several available encoded channels (permitting, for example, switching to a different resolution during a connection).

Since several simultaneous coded versions need to be generated from a common video signal source, this approach is only appropriate for real-time applications. It is inappropriate for stored video applications, since it would require wasteful storage of multiple versions of the same signal (and possibly duplicated processing such as mixing, inserted windows, etc.) or transcoding upon reading the signal from storage to convert it to multiple coded formats.

Layered Signal Approach:

(This is called the "embedded bitstream" approach in CMTT/2-18. See AVC-28).

A hierarchical representation of the video signal is defined. Coders transmit a baseband signal which provides a basic quality service. Incremental signals, which can be used along with the baseband to recover a high quality signal, are also transmitted. Receiving terminals utilise the baseband and an appropriate number of incremental signals to recover the video signal to the quality which they are capable of displaying. Transmitting terminals provide the number of signals which is commensurate with their input signal quality. Layered signal coding has the interworking flexibility of simulcast, but the use of incremental signals should lead to improved channel utilisation. It is a suitable coding method for communicative, storage and distribution applications. Furthermore, layered coding is a likely candidate for video coding on ATM networks independent of the question of service integration, due to its inherent cell loss tolerance. Future enhanced services can be provided by transmitting additional incremental signals without affecting lower layer services.

Comparison.

The following table provides a comparison of the methods of video service integration discussed above:

Interworking method:	Negotiation	Simulcast (single decoder)	Simulcast (Mult. decoders)	(Fully) Layered*
Rate overhead	Low	High	High	Low-Medium†
Complexity	Medium	Low-Medium	High	High
Inherent cell loss tolerance	No	No	No	Yes
Suitability for Broadcast/multipoint	Poor	Good	Good	Good
Suitability for stored video	Poor	Poor	Poor	Good
Upward compatibility+	Yes	Yes	Yes	Yes
Downward compatibility#	No	Yes	Yes	Yes
Change of quality level during connection	Yes	No	Yes	Yes

Table AVC-35/1. Comparison of approaches to video service integration.

The comparison table shows that end-to-end negotiation of parameters is not a suitable means of providing service integration, if for no other reason than its inability to accommodate point-to-multipoint connections with a range of different terminal types. Simulcast is more flexible, but it would incur a significant coding rate overhead, could lead to network overheads in establishing multiple simultaneous connections and it lacks the inherent cell loss tolerance of the layered systems. Importantly, simulcast also would appear to be unsuitable for use with stored video applications.

---

* The term "fully layered" is here intended to represent a system in which any transmission involves layers from the lowest defined quality level upwards. A more flexible system is introduced in Section 4.

† The transmission rate overhead for simulcast transmission will be of the order of 30-40% compared with a single, one-layer channel. The overhead for a layered system is likely to be somewhat less than this, but the actual penalty requires further investigation. Note that layering may be required for cell loss protection anyway, and the relative cost of providing service layering may be small.

+ Upward compatibility is defined in CMTT/2-18 (see AVC-28) as the ability of a high resolution receiver to decode signals from a low resolution encoder. It is questionable whether the word "compatibility" is appropriate in the case of simulcast systems, since the new service class is introduced essentially independent of the old encoders and decoders.

Downward compatibility is defined in CMTT/2-18 (see AVC-28) as the ability of a low resolution receiver to decode signals from a high resolution encoder.

The question of cell loss tolerance is significant since the liaison from SGXVIII indicates that cell losses, when they occur, are likely to occur in bursts (see AVC-24). While forward error correction techniques may be tolerant of isolated cell losses, it is only layered coding systems that are likely to be tolerant of cell loss bursts (in the low priority channel). Clearly the layered coding approach to interworking has many advantages, including coverage of all service applications, but it also has some disadvantages, as discussed in Section 3.1.

Assuming that a multi-service or multi-purpose terminal will have a single common output buffer and display, there is a need to ensure that the set of formats that can commonly be expected will be compatible with this single output stage. It is important to note that <u>any</u> system offering integration of video services relies on the definition of an appropriate family of picture formats. This important issue is addressed further in AVC-36.

3.1. Potential Problems with a Fully Layered System

Although the above comparison indicates that a layered or hierarchical signal structure is the most attractive in offering the advantages of services integration, and it is inherently well suited to ATM transmission due to its tolerance to cell loss, a fully layered system does have some potential drawbacks. Layered coding (for interworking) is normally depicted as in Figure AVC-35/1 below, where any input video signal is decomposed into a common base layer and multiple additional layers to allow definition of the signal at several quality levels up to and including that of the original source*. In terms of some common services that could be expected on the B-ISDN, these might correspond to videophone, videoconference, standard television and high definition television qualities (resolutions, frame rates, etc.).

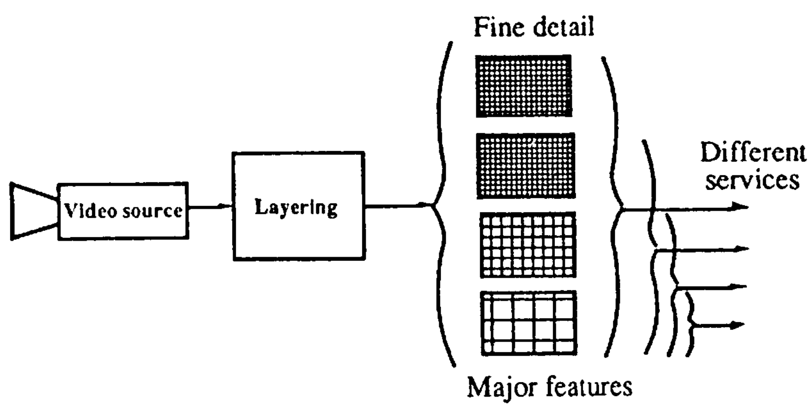

Figure AVC-35/1. Full layering, in which a common base layer is used for every video service.

While satisfying the goals of integration, such a structure is open to criticism on several grounds:

- <u>Complexity</u>. The need to decode perhaps four layers at HDTV level and combine the various layers to regenerate a single signal could be seen as excessively complex.
- <u>Cost</u>. Such increased complexity could lead to increased cost. This could be particularly important when a new service is introduced, even though the interworking capability of the layered system would assist in the introduction phase.
- <u>Coding penalty</u>. While the overheads involved in layered coding are yet to be quantified, it seems likely that there will be some, compared with single-layer coders. (As noted earlier, there may be little coding penalty compared with a layered coding system providing cell loss robustness.)
- <u>Permanency of lowest layer</u>. It could be argued that, with evolution of the various video services, and decreasing transmission costs, lower grade services may disappear. For

---

* This corresponds to the "embedded bitstream" method described in CMTT/2-18 (see AVC-28).

example, QCIF resolution for videophones may be replaced by CIF. The structure of Figure AVC-35/1 would mean that all services would be forced to continue performing unnecessary processing at the lowest resolution level.

- Unknown usage expectations. While the full layering structure offers maximum flexibility (e.g. viewing of an HDTV-originating picture on a videophone receiver), it is unknown whether such a capability will be necessary/desirable. It may be demonstrated to be unnecessary in practice.

An approach that offers the benefits of fully layered coding, while avoiding the above penalties, is desired.

4. Proposal: Flexible Layering

The penalties associated with fully layered coding systems can be avoided if we remove the constraint that all video signals must use a common base level, and instead permit flexibility to introduce as many layers (going down from the input quality level) as desired. In this case, any quality layer could provide the base signal, with as many incremental levels above it as desired. This is illustrated in Figure AVC-35/2 with some examples.

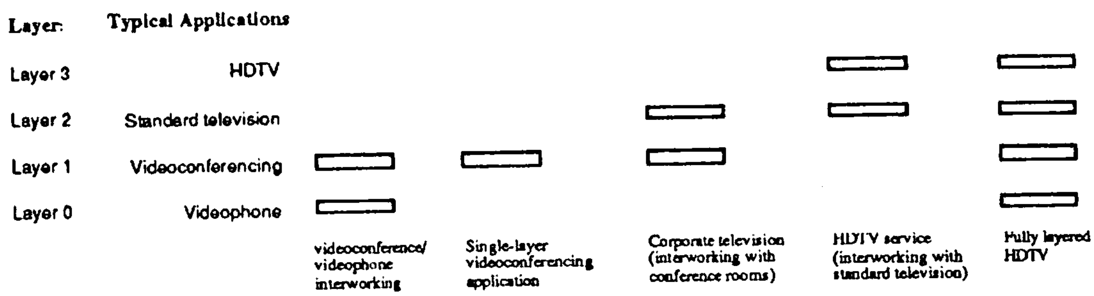

Figure AVC-35/2. Examples of layered structures that could be accommodated with flexible layering. Note that a single layer coder is accommodated as a special case.

The flexibility of the system would permit additional higher layers to be introduced progressively when the costs and benefits warrant them. The layers used during a connection could also be configured at call establishment, and could be added or dropped during a call to economise on network costs. (Note that additional lower layers could not be introduced if the receiving terminals only implement upper-layer processing. This is not likely to be a long term problem, since the multiservice, multimedia nature of future communications terminals will encourage the incorporation of processing capabilities down to the lowest possible layer.)

Examples where this system might be attractive are:

- a two-layer TV/HDTV system could provide interworking between these two service categories, without the need to provide for lower quality interworking if this is considered unnecessary.
- a videoconference level video system may operate using a single layer system if the probability of cell loss is sufficiently small that the visible degradation is considered acceptable in a particular configuration. Such a configuration would have minimum coding overheads and complexity.

This second example illustrates a very important point; that the proposed flexible layering architecture subsumes single layer coding as a special case. Therefore, acceptance of flexible layering does not preclude other approaches. The important decision, however, is to adopt a layered framework so that both multi- and single-layered coding schemes can be used and interwork as much as possible. Restrictions on the architecture are few:

- a suitable set of picture parameters defining the levels must be determined (see AVC-36).
- there will be a requirement for end-to-end signalling to indicate the layers used, and any coding changes (e.g. different quantisers depending on whether a layer is a baseband or incremental signal).

It is, of course, important to determine that the flexible layering approach is possible. It is noted here that at least one architecture, a pyramid-based coding system, can provide this capability. In this approach, subtractions are performed to generate incremental signals, giving the advantage that coding and decomposition at all layers can be independent and therefore optimised for the particular service category. This and other candidate flexible layering architectures are currently being studied in Australia.

5. Conclusion

To promote video service integration and gain the full benefits of interworking between video services on the B-ISDN, a layered framework has been proposed. Flexible layering gives complete freedom to use only as many layers as is necessary, practical or economical for a given service while still conforming to the basic framework. It is proposed that the Experts Group adopt this framework and encourage its use for all video services on the B-ISDN by proposing it to CMTT and other interested parties through liaison statements and the IVS Baseline Document.