Telecommunications Standardisation Sector
Study Group 15
Experts Group for ATM Video Coding
(Rapporteur's Group on Part of Q.2/15)

Document AVC-559
2 September 1993

Source: BT
Title: Comments on Bellcore continuous presence multipoint proposal.

Background

This contribution is in response to Documents AVC-500 and AVC-501 from Bellcore which propose a method of achieving 2 x 2 split-screen continuous presence multipoint videoconferencing.

The proposal makes use of a known feature, intentionally included when writing Recommendation H.261, to permit in principle the mixing of 4 QCIF pictures into 1 CIF picture by manipulating only the picture header and group of block (GOB) header information, without actual video processing. By avoiding the need to variable length decode and reconstruct the images from each terminal the method reduces hardware costs and potentially reduces end to end delays in comparison to a full decode and mix in the pel domain.

Discussion

In the Bellcore proposal, terminals will have to transmit QCIF pictures using no more than one quarter of the available bit rate and must be able to receive a CIF picture at the full bit rate. In most networked applications the links to an MCU have symmetrical transmit and receive data rate. In the standard H.221 multiplex the amount of data allocated to video is normally symmetrical. Therefore if the combination of the 4 QCIF pictures must fit into the available bit rate the independently produced QCIF pictures can only be allocated one quarter of the available bit rate. (Very slightly more is possible because of the removal of the picture headers from 3 of the 4 inputs).

It would be very difficult to satisfy the need for the QCIF pictures to run at one quarter of the CIF rate by defining new bit allocations for video in the H.221 multiplex as the video data normally fills all bit positions not occupied by other services. The proposal therefore suggests that the H.261 error corrector fill bit mechanism should be used to generate a video bit stream in which only one quarter of the bits are active video.

Considering some topics in detail:

1. Although the use of the fill bit requires no changes to H.261 as a standard it does constitute a significant change in the way that existing H.261 terminals use the fill bit. Normally the fill bit is only asserted when buffer under flow is imminent. For this reason and for others discussed later in this document it is not possible to implement this proposal with existing terminals without modification to those terminals. It is likely

that some existing terminals and H.261 chipsets will not be capable of using the fill bit to deliberately reduce the active video rate.

2. When participating in a conference it is very helpful to be able to see all the participating sites, but in many situations conversation is carried on between two parties with other parties listening in. During the conference the dominant parties may change and at times several parties may contribute, but not all the time. Ideally, each site would be able to see all other sites to have a feeling of their presence, whilst also seeing the current speaker in more detail. So for example the current speaker would appear as a CIF image on one screen and up to 4 locations would appear combined in 1 CIF picture on a second screen.

With the Bellcore proposal only a low bit rate QCIF picture is available from each terminal. For a 2B conference the video rate, assuming 48 kbit/s G.722 audio, no data and no encryption, is 78.4 kbit/s. Each individual QCIF rate will be limited to 19.6 kbit/s. The display of this picture as a full size image on anything other than a small screen would not be acceptable. (This can be confirmed by pictures seen at the meetings of the Rapporteur for Very Low Bit Rate Coding.)

3. The simulation studies paper (AVC-501) concedes that although the pel-domain combining results in slightly degraded picture quality, the degradation is insignificant and almost undetectable subjectively. This conclusion is based on comparing the picture quality before and after a decoded picture has been re-coded and decoded a second time. A more relevant comparison might be between the following two cases. The first is a double encode and decode entailing an initial encoding and decoding using CIF at the full bit rate available from a terminal followed by coding and decoding with QCIF at one quarter of the bit rate. The second case is the single encode and decode with QCIF and quarter bit rate which corresponds to the Bellcore proposal. It is expected that the objective and subjective degradations introduced by the double encoding will be significantly less than those observed in the Bellcore experiments.

4. The simulation studies paper demonstrates end to end delays for a 256 kbit/s video bit-rate (64 kbit/s per QCIF source) of between 266.67 and 644.44 ms. The mean delay appears to be approximately 440 ms. At 256 kbit/s the end to end delay for a single CIF encode and decode is approximately 260 ms and does not vary much with picture content. This gives a back to back delay of 520 ms without any optimisation. The proposed method of combination therefore appears to give only a marginal improvement in end to end delay and in some cases causes a degradation.

5. The hardware to support the combination of pictures at the GOB level may not be as complex as that required to decode the picture and re-code it but all of the complexity saving lies in the compression process for which chipsets and complete codecs are readily available. (Line terminating, clock extraction, data extraction etc are necessary for both approaches.) The cost of an H.261 encoder and decoder is decreasing all the time, and soon its marginal cost over finding and manipulating the picture and GOB headers and controlling the input buffers will be insigificant in the overall consideration.

It is arguably a more complex task to build a 4 QCIF to CIF combiner than to put a decoders back to back with an encoder through a 4:1 spatial decimator/combiner.

6. It is not clear whether the simulations performed have taken into account the dropping of frames by the source encoders. At low bit rates it is quite normal for an H.261 encoder to drop frames rather than quantise a picture more heavily in order to limit the coded bit-rate. The dropping of frames results in those frames that are transmitted being allocated more bits than average. The effect of this will be similar to the Intra coded frame in response to a fast-update request (FUR). Unlike the FUR the timing of the dropped and transmitted frames can not be controlled. It is quite likely that all 4 QCIF sources will drop frames at the same time and then start to code pictures with a large number of bits in phase. This has been shown to result in longer end to end delays. It is questionable whether there would be much gain in distributing the FURs in this case.

7. The dropping of frames also has an implication on the way the 4 QCIF pictures are combined. The Bellcore proposal suggests that the Temporal References from all but 1 source are ignored and that pictures are combined as they arrive to achieve minimal delay. With dropped frames this could result in temporal distortion.

8. Document AVC-501. section 4.2 states:

"The videoconferencing terminal clocks can be locked to the network clock so that data transmission is synchronous between the terminals and the network. These synchronous terminal clocks also ensure the frame rates of different QCIFs are exactly the same but their frame phases may be different."

Such locking is not common-place in terminal designs. BT is not aware of any which incorporate it. While it is technically feasible, it does place restrictions on the selection or design of cameras. For integrated terminals these may be under the control of the designer but for additional plug in cameras and other sources such as video tape playback machines, the locking requirement is a drawback.

9. The proposal requires the introduction of new capability and command codes that must be transmitted and understood by the video terminals. This precludes the use of existing terminals without an update, if possible, to their H.242/H.230 software. As mentioned above the terminals also have to be modified to use the error corrector fill bit for quarter rate video.

CONCLUSION

Though Documents AVC-500 and 501 are most welcome as representing needed and worthwhile investigations into the practical usefulness of the untested hook intentionally inserted in Recommendation H.261 some five years ago, BT is of the opinion that the technique is not yet proven and may always suffer from disadvantages.

It has not been shown that the pel domain combining method has sufficient subjective performance drawbacks to negate its very real advantages of:

1. needing no changes to existing Recommendations,
2. working with existing terminals and
3. being much more flexible. As examples:

- It can provide both switched and continuous presence multipoint, even within one conference. Each receiver could (statically or dynamically) notify the MCU whether it wished to receive a combination of the pictures from the other locations or the higher quality signal from just one other, such as the current speaker.

- An MCU which decodes, combines in the pel domain and recodes is able to adapt the effective portion of the bit rate given to the pictures from each site. It can allocate more bits dynamically to the more active locations. It is also able to provide a 4 location meeting with pictures from the 3 other sites occupying 3 of the 4 quarters of the CIF composite but each getting a third of the bit rate.

- The pel domain combining type of MCU is able to establish a multipoint connection between CIF, QCIF and even PSTN videophone terminals, with each connection being at an independent bit rate and without all being brought down to the lowest performance.

End