CCITT SGXV
Working Party XV/1
Specialists Group on Coding for Visual Telephony

Document #279
Jan. 26, 1988

SOURCE: Japan
TITLE: Proposal of mx64kb/s Codec Picture Format

1. Introduction

Application of mx64kb/s codecs will be extended in the future to various audiovisual services including videophone and videoconferencing. As for terminal equipment configuration, not only dedicated communication terminals but also PCs and WSs with videophone terminal capability may be manufactured.

On the other hand, most users expect a reconstructed image quality close to that of today's broadcasting television, or equivalently CIF* quality, regardless of the communication channel bit rate. However, coding efficiency obtained now and to be obtained in the near future (1990-1992) may not be enough to get up to CIF quality at 64kb/s. In fact, a codec designer has to choose where to place priority, on temporal or spatial resolution, according to the application and his designing philosophy. For example, a high coded frame rate with a medium spatial resolution may be a guideline for designing videophones, whereas videoconferencings with interactive graphics may need high spatial resolution with less coded frame rate. If techniques to get CIF quality are developed, the cost to realize them will much increase.

* It is noticed that 'CIF' appearing in this document includes the temporal resolution specification.

Considering these prospects, it can be concluded that a fixed picture format is not applicable to all the required services from technical view point because the required characteristics change in terms of the application.

This document proposes a variable picture format in order that the format can be applied for all needs or services and it can evolve with progress of coding techniques. The proposal is summarized as follows:

1) The aspect ratio of a pixel is identical to that of CIF.

2) The maximum format is 352 pels, 288 lines per picture with 1/29.97 sec of minimum coded frames interval (same as CIF).

3) One can use formats smaller than the maximum format (CIF) in a session. The variable ranges are as follows:

   horizontal pels: 32 x k, 5=<k=<11 (160 to 352 pels)
   vertical lines: 16 x l, 8=<l=<18 (128 to 288 lines)
   minimum coded frame interval: f/29.97 sec, 1=<f=<5-10**

   ** Exact value is for further study.

4) In order to ensure connectivity among all codecs and minimum quality, the following two points are proposed.

a: Minimum capability specification. All coders and decoders must have a capability to handle a format not smaller than (k, l)=(8, 12) and f=2 (when k, l=8, 12).

b: Specification for connectivity of all codecs. All coders and decoders must be able to handle the all formats not larger than the maximum format they have.

5) The format used in a session is determined by negotiation process.

In section 2, the variable format system is defined. Section 3 specifies minimum facilities of coder and decoder to ensure connectivity and quality. A guideline of picture format negotiation is described in section 4. Section 5 mentions compatibility with nx384k/s codec. In section 6, several considerations are noted.

2. Parameters of variable format

The numbers of horizontal pels and vertical lines for luminance are represented as integer multiples of 32 and 16, respectively, while the minimum coded frame interval is represented as a multiple of 1/29.97 sec.

1) Luminance: 32 x k, 5=<k=<11, pels in horizontal (160 - 352 pels)
16 x l, 8=<l=<18, lines in vertical (128 - 288 lines)
2) Unit of change: 32 pels and 16 lines
3) Minimum coded frame interval: f/29.97 sec, 1=<f=<5-10
4) Pixel aspect ratio: 4/360 : 3/288 = 1.0666666.. : 1.0
5) Subsample rate for chrominance: horizontal 1/2, vertical 1/2

A format can be represented by 'k', 'l' and 'f'. For example, CIF corresponds to (k, l, f)=(11, 18, 1).

3. Minimum capability of coder and decoder

In order to ensure the connectivity and quality of all codecs, following points are mandatory for all codecs.

1) All coders and decoders must have a capability to handle a format not smaller than (k, l)=(8, 12). In other words, maximum available picture size of a coder or decoder must exceed or equal to 256 pels and 192 lines.

2) All decoders must be able to decode a signal with coded frame interval of 2/29.97 sec when the format is (k, l)=(8, 12).

3) All coders and decoders must be able to handle all formats not larger than the maximum format they have. For example, a decoder with maximum available format (k, l)=(9, 15) must be able to handle all formats of 5=<k=<9 and 8=<l=<15.

4. Format determination in a session

Picture format parameters k, l and f are determined at the start of a session, or possibly during a session, by negotiation through Application Channel etc (exact negotiation channel is for further study waiting for the study on audiovisual intercommunication protocols). The negotiation consists of following two steps.

1) Capability exchange. After this sequence, both sides know the maximum format available in the session.

2) Actual mode determination. An actual format used in the session is determined through a negotiation process. Picture size is first determined, then minimum coded frame interval is negotiated.

5. Compatibility with nx384k/s codec

When picture format parameters k, l and f are equal to 11, 18 and 1, the format is equivalent to CIF. In that case, compatibility between picture formats of 64kb/s and 384kb/s are ensured. Even if k, l and f are not 11, 18 and 1, simple relation to CIF is still maintained because the pixel aspect ratio is common.

6. Discussion

1) Aspect ratio of image

Proposed variable format system allows pictures with horizontally wider or vertically taller aspect ratio than normal television. This flexibility increases applicability of codecs to various needs and services. For example, videoconferencing with three persons requires an area shown in Fig. A2 (c) and personal videophone requires an area shown in Fig. A2 (e). Coding area limitation saves unnecessary bits and give higher reproduced image quality. Various informations, e.g. telephone number of the opposite side, can be displayed outside the coded window. In some multipoints realization, such display modes as in Fig. A2 (a), (b), (d) and (f) can be used according to the number of terminals and attendances.

2) Format conversion

Format conversion process between local television and the common digital picture format is not a part of recommendation. The processing methods can be classified as follows:

a: TV <--> CIF conversion is performed at the input of the coder and the output of the decoder. A necessary area is cut out from CIF. In this case, full scan screen is obtained only when the format is same as CIF.

b: Television signal is first sampled at 13.5MHz, then pixel density conversion is performed by digital filtering. If the format is 256 pels and 192 lines for example, full scan screen is obtained by the following process.
   i) The sampling frequency is 13.5MHz, 720 active pels are obtained.
   ii) 720 (active) + 48 (inactive) pels are reduced to 256 pels by 1/3 subsampling.
   iii) 4/5 (525 region), 2/3 (625 region) line density changing.

c: Other methods where sampling frequency different from 13.5MHz is

used.

When coded window is smaller than the area taken by camera, visual feedback through picture monitoring function to identify the coded window is supposed to be necessary.

3) Hardware

Generally speaking, impact of picture format to hardware quantity can be evaluated in two categories. One is the hardware quantity to get required picture format from input (analog) signals. In the format conversion process, it may be dominant for hardware quantity whether the number of vertical lines of the format exceeds that of a field or not until format conversion VLSI becomes available. The other is the coding or decoding processing quantity, which increases as the number of processed pels in a unit time increases.

In the proposed variable format, additional hardware for changing coded windows (the changing range is 5,8 =< k,l =< maximum capability of the codec) is required. However, the variable format has an advantage that service or equipment providers can choose where to place features on their codecs, e.g. maximum capability of temporal and/or spatial resolutions, auxiliary information display function, multipoint display function, cost, etc.

This section gives a prospect of hardware increase by adding variability of coded window:

(a) Format conversion process:

Format conversion process between analog signal and the common digital picture format is not a matter of standardization. If one designs the format conversion to display or take fully scanned image irrespective of the picture size, he needs much hardware* for pixel and line densities conversion. In case that under scanned images are allowed, smaller picture sizes are obtained by cutting out from the codec's maximum picture size. Additionally required hardware for the cutting out is estimated very small.

* If the possible number of vertical lines is limited and DSP (or other programmable device) is used for changing horizontal pixel density, the hardware quantity may not increase so much.

(b) Coding process:

In wired-logic (or ASIC) approach, the variability of coded window can be realized with not so much additional hardware, if the hardware is designed to include the variability from the start of designing. In DSP approach, variability of coded window may be realized with little additional hardware. Another merit of employing DSP approach with variable picture format is that one can design codecs to improve the coded frame rate as the number of pels in a frame decreases.

4) Unit of picture size change

It has been reported that we can reduce overhead bits by transmitting the block attribute information for a larger block such as 16 X 16 instead of for a block of 8 X 8. When utilizing this fact (Fig. 1), 32 pels and 16 lines are appropriate for unit of picture size change. (Chrominance signal is 1/2 subsampled.)

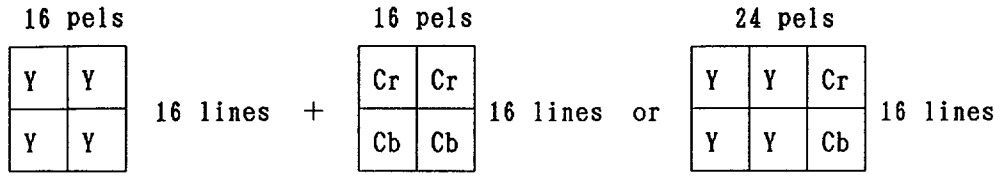


Figure 1. Example of larger block size use for block attribute transmission

5) Minimum capability

The minimum capability (256 pels, 192 lines and 2/29.97 coded frame interval) is selected as a format giving minimum tolerable quality to display three parsons seated side by side in videoconferencing, and having nearly 4:3 aspect ratio.

6) Number of picture format

The number of possible picture formats may be restricted as depicted in Appendix 1 considering possible application and hardware complexity. In order to reduce hardware complexity, it is preferable that the possible number of vertical lines is limited to a small number.

7) Problems in case of employing full CIF

If full CIF is employed as the only one mx64kb/s codec picture format, the following problems rise.

a: In the near future, coding algorithm performance is not supposed (but expected) to reach full CIF level. Therefore, when the first version equipment following the recommendation is put into market, the codec may have sufficient spatial resolution but limited temporal resolution. This situation is not acceptable for the people who believe temporal resolution is important for natural reproduction of videophone pictures.

b: All decoders must have capability to decode 352 X 288 image in every 1/29.97 second. This requirement needs more hardware.

8) VCR demonstration (Appendix 2)

7. Conclusion

The variable picture format for mx64kb/s codecs has been proposed.

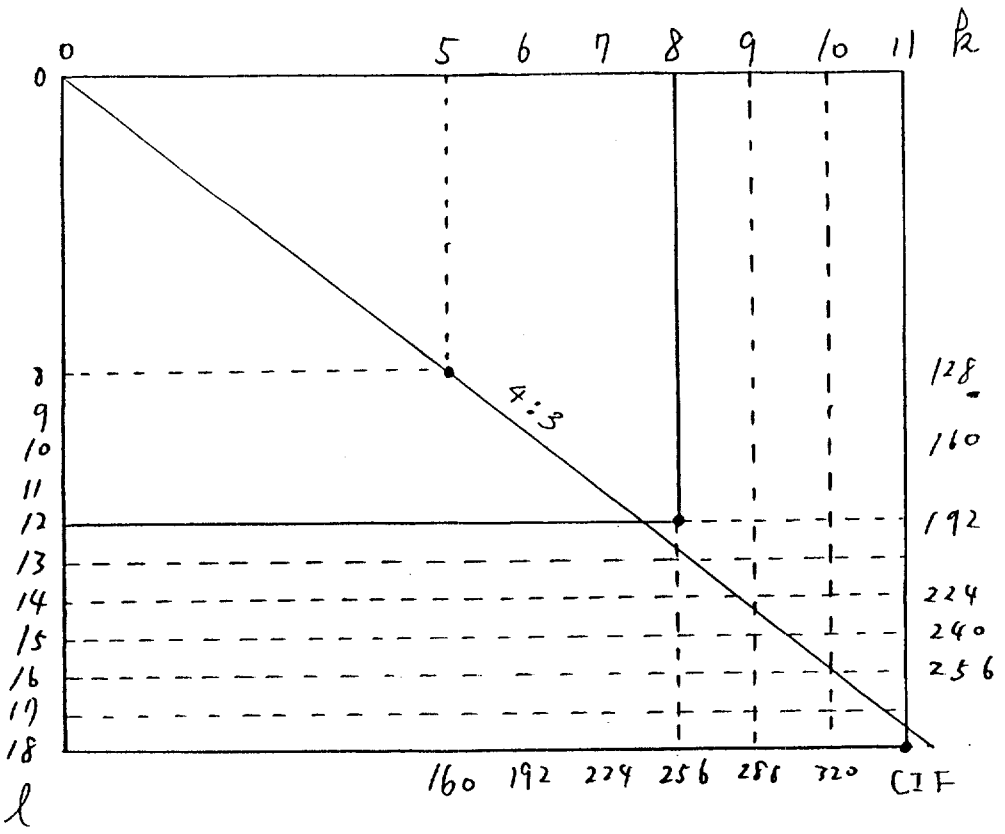


Figure. Variable Format

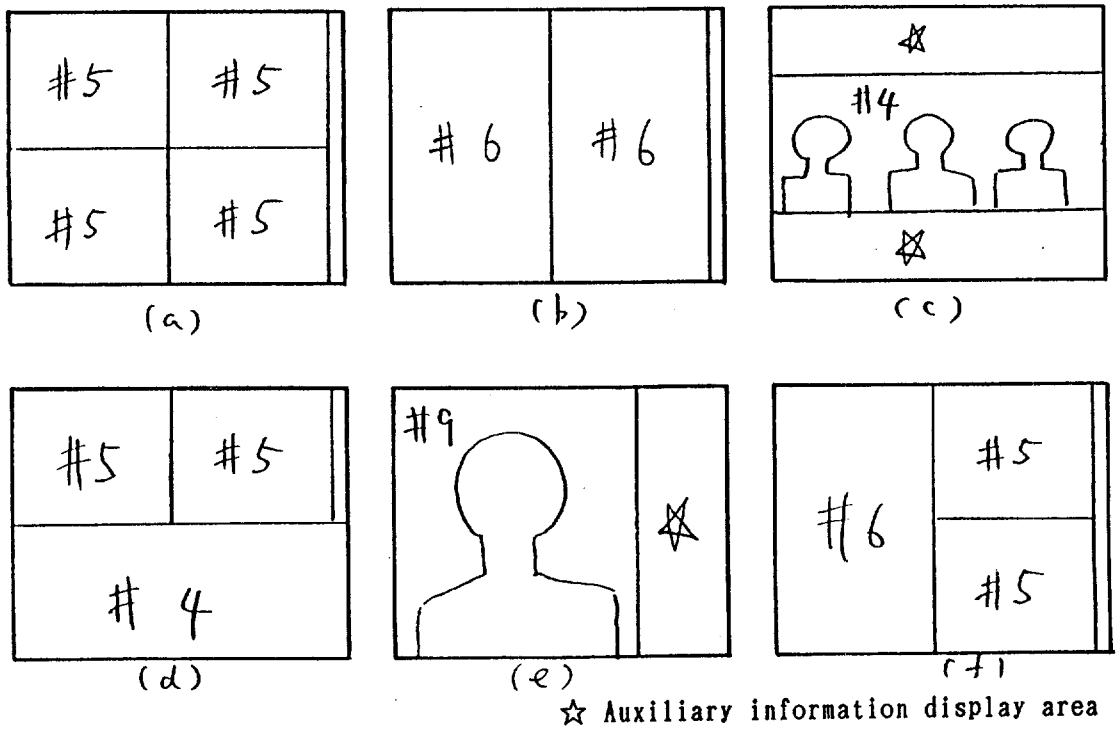
59

The number of possible picture sizes (possible combinations of k and l) is theoretically 77 in the proposal. We can reduce the number of (k, l) combinations to 10-20 like Figure A1 for example. Here only the combinations with #number are possible. Using picture formats in Figure A1, practically useful display modes shown in Figure A2 are possible for example.

k / l	5	6	7	8	9	10	11	
1								
8								128
9	#5			#7			#4	144
10								160
11								176
12				#2				192
13								208
14								224
15	#8			#3				240
16								256
17								272
18	#6			#9			#1	288
	160	192	224	256	288	320	352	

#1 CIF.
#2 Minimum Capability.
#3 256 pels 240 lines.
#4 For videoconferencing.
#5 See Figure A2.
#6 See Figure A2.
#7 for videophone. (two people)
#8 for vidoephone.
#9 for videophone. (Fig.A2)

Figure A1. Example limitation of (k, l) combinations



☆ Auxiliary information display area

Figure A2. Application (example)

VCR Demo.

Part 1. Resolution changed and under-scanned image examples.

No 1. Original CIF
No 2. Resolution changed (5/6)
No 3. Resolution changed (first 5/6 then 6/5)
No 4. Resolution changed (2/3)
No 5. Resolution changed (first 2/3 then 3/2)
No 6. Cut out (5/6)
No 7. Cut out (2/3)

Part 2. Application examples