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

Document # 60

January 13, 1985

SOURCE: NTT, KDD, NEC and FUJITSU

TITLE : GENERIC STRUCTURE OF 384 x n kbit/s CODEC

---

1. Introduction

At the last meeting in Torino, several candidate algorithms were presented and a general view was obtained that future study should focus on the scheme with motion compensated interframe prediction plus further processing (see 4.1(2)/Document #54R). According to this guideline, study on DCT-based and VQ(Vector Quantization)-based algorithms has been carried out in Japan.

This document describes a generic structure covering both algorithms, paying attention to the information to be transmitted. Common parameters for coding algorithm simulation work are also presented.

2. CODEC Structure

2.1 Outline

Based on the fundamental configuration shown in Figure 1/#43, a more detailed version is elaborated as Figure 1. One of the essential features is that the video coding loop, which processes the motion-compensated interframe prediction errors, consists of 'coarse(C)' component coder and 'fine(F)' component coder. C-component has a nature of low-passed, envelope or more frequent signals, while F-component has a nature of high-passed, pulse or less frequent signals. In other words, C-component corresponds to a core part and F-component corresponds to an enhanced part.

In order to get improved coding efficiency required for sub-rate codec, C-coder should operate on a block basis even if there would be a risk of ~~bearing inherent mosaic pattern or blurred edge distortion. The~~ following two alternatives are considered.

a. DCT-based processing in the coefficient space
b. VQ-based processing in the pel space

As for F-coder, it should operate in pel space to process almost uncorrelated components.

2.2 Brief Description of Each Functional Block

2.2.1 Television Standards Converter (TSC)

Conversion to/from the agreed intermediate format takes place. The conversion algorithm follows Document #55 in principle.

2.2.2 Pre-filter (PrF) and Post-filter (PoF)

Temporal and/or spatial filtering to reduce random noise and distortion included in the original input signal and coded output signal. In hardware, this block may be incorporated in TSC.

When frame dropping is used in the coder, motion compensated interpolation may be required to reduce jerkiness in the post-filter.

2.2.3 Motion Detector (MD), Variable Delay (VD) and Frame Memory (F)

These three functional blocks carry out motion-compensated interframe prediction. Intra-frame predictor(s) for scene change operation may be included.

2.2.4 Threshold (THL)

This identifies insignificant picture blocks whose power is negligible. Hence, all of the error pels contained in the block can be treated as zero.

2.2.5 C-coder ('C' COD)

Main target for the standardization study. Details of experimented encoding and quantization are discussed in the companion documents (#61, #62).

2.2.6 F-coder ('F' COD)

Amplitudes and positions of significant pels are quantized.

F-coder has the following role in each C-component coding method. For DCT-based coding, F-component is impulsive or spatially high-passed signal which can not be efficiently coded by transformation. It works as edge enhancing to the moving regions.

In VQ-based coding, the necessity of F-coder comes from a restriction of realizable number of vectors in the codebook. Some blocks can not be matched with any of the representative vectors. Residual error should be recovered by F-coder.

2.2.7 Shaping Filter (ShF)

Spatial filter to remove granular noise.

2.2.8 Variable Length Coder (VLC)

Data compression using statistical characteristics of each unit of information to be transmitted.

2.2.9 Transmission Buffer (TB)

Irregularly spaced bursts of coded bits are converted into a uniform rate bit stream.

2.2.10 Control (CONT)

Using the buffer memory occupancy and other statuses, coding and quantizing parameters are controlled to keep the generated information quantity in an appropriate range. Both feed-forward and feedback should be utilized.

If a scene change is detected by, for example, the amount of information generated in a frame, one of the following methods can be employed.

- To freeze the frame just following/preceding the scene change and transmit the new frame during subsequent several frame period.
- To freeze the frame just preceding the scene change and wipes it out with a new scene by increasing the coding area frame by frame.

2.2.11 Transmission Coder (T COD)

Error correction, encryption, multiplication of coded audio, coded video, data and control signals into a frame structure are carried out.

2.3 Information to Be Transmitted

The video source coder having a generic structure mentioned above outputs the following four different kinds of information, which are transmitted to the decoder through the transmission channel.

a. Motion vector for each block of motion compensation (v)
b. Flag to identify whether a block to be coded is significant or insignificant (t)
c. C-component quantization index (c)
d. F-component quantization index (f)

It should be noted that all of the above include some address information showing the position of the block, pel or coefficient concerned.

3. Parameters for Simulation Work

A fair number of rules for algorithm comparison were established in Torino (see 4.2/#54R). Following these rules, simulation work has been carried out. Common parameters used for simulation of DCT-based and VQ-based algorithms are described herein. The parameter values are provisional for the purpose of having the same ground in different organizations. Optimization will be further studied.

3.1 TDM Format

Of the three conceivable formats shown in Figure 2, the first one is employed for picture coding efficiency. The second one might be more appropriate for block-line (e.g. 8 lines in case of 8 x 8 block size) basis coding mode control, prevention of adverse effects due to possible coding stop in the middle of the picture, delay and hardware simplicity.

3.2 Pre-filter

A temporal filter with nonlinear characteristics (Figure 3) is used.

3.3 Motion Detection

1) Detection method : block matching
2) Block size : 8 pels x 8 lines
3) Tracking range : +15 pels x +11 lines
   Movements of videoconference participants have been considered.
4) Possible output vectors: all of the vectors included in the above tracking range
5) Evaluation measure : summation of absolute matching error as regards the 64 pels in a block
   In a real hardware, performing all of the 64 calculations will not be easy. Some subsampling may be required.

3.4 Identification of Significant Blocks

Total power (summation of squared prediction error value) in a block determines whether the block is significant or not.

3.5 Shaping Filter

Though this filter is concerned with compatibility, there is not sufficient information at the moment. If this filter is used in simulation, its characteristics will be described.

3.6 F-coder

Amplitude is scalar quantized, and the position of non-zero pel is transmitted using run-length coding. Method of scanning is open because it probably does not affect coding efficiency.

3.7 Variable Length Coding

For estimation of coded bits, Part 3 CODEC tables are applied. Non-zero elements, whether they are pels or transform coefficients, are coded using Table 5/Document #19, while zero elements are coded using Table 6/#19. As for motion vectors, Table 7/#19 is applied with a minor extension in vertical motion size and length of the run.

3.8 Transmission Buffer

Since the total delay between encoder and decoder is envisaged as around 200 ms, and television standards conversion to/from the intermediate format may require about 50 ms, a buffer size of 50 kbit is selected. It has been considered that better picture quality can be expected with a bigger buffer size. Definition of delay time due to occasional field dropping needs further study.

3.9 Post-filter

When the number of coded frames are 15 per second, omitted frames are simply repeated. Motion compensated interpolation seems indispensable if more frames are omitted. This point is for further study.

4. Items to Be considered for Algorithm Selection

The following items should be considered to finalize the 384 x n kbit/s coding algorithm to be recommended.

a. Picture quality
b. Hardware implementation
c. Simplicity of compatibility related specifications
d. Ease of coding control
   - Adaptation to various input pictures
   - Adaptation to various output rates (e.g. value of n)

5. Some Guidelines to Further Specify the Coding Algorithm

1) When conditional replenishment is employed, either of the following problems arises:
- Still backgrounds tend to fluctuate because conditional replenishment works to renew them.
- Dirtied backgrounds remain without being renewed.

Some method of 'cleaning up' will be required, which is activated when the amount of coded picture information is small.

2) The number of different kinds of transmitted information is required to be as small as possible, from algorithm description and hardware size

points of view. If multiple kinds of information are transmitted, care should be paid to preventing multiple coding of the same information.

3) When coding mode control on a block-by-block basis is used, attention should be paid to the effects on picture quality and additional overhead control information.

4) In order to confirm the validity of a specific coding algorithm, stress materials which seem difficult to process, e.g. a much detailed picture, should be tested at an early stage in addition to the agreed sequences. No catastrophic phenomena should be allowed.

5) If large quantization noise arises in the color difference signal, a reproduced picture is susceptible to dirty window. Pre-emphasis or an equivalent operation is desirable.

6) In simulation, buffer occupancy is approximated by the amount of information generated in a block-line, for example. The delay from the coding loop output to buffer memory input should be taken into account so as not to cause buffer overflow.

6. Conclusion

A generic structure and common parameters for the two coding schemes, DCT-based and VQ-based, have been described. Details of each coding algorithm are presented in companion papers. Early decision of the coding algorithm based on the items listed in 4. above is proposed toward the target of making a recommendation at the end of this study period. For that purpose, only compatibility related items should be specified leaving the possibility of introducing future advances.

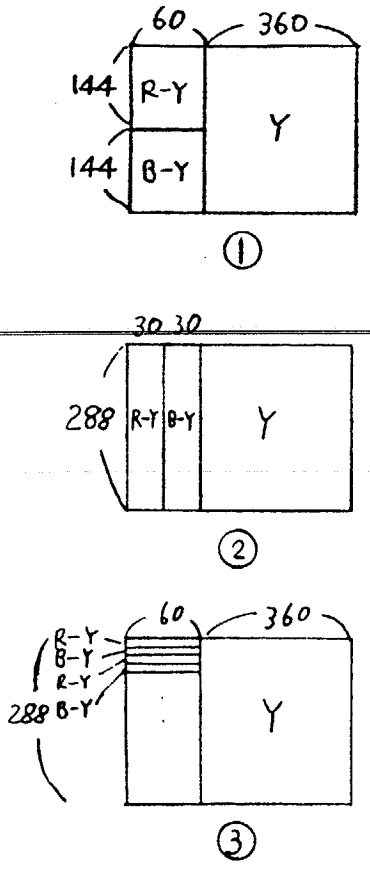


Figure 2 TDM Format

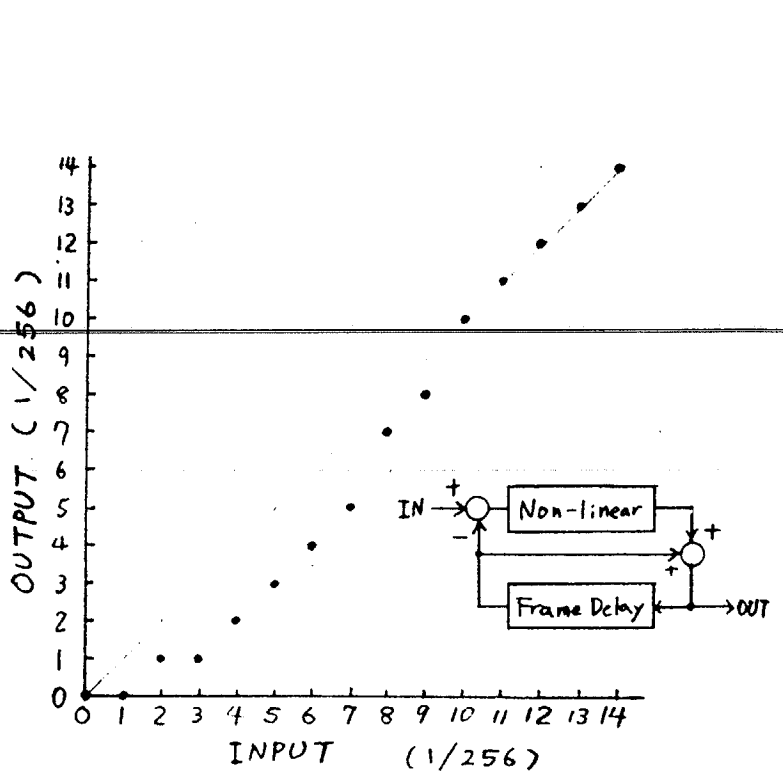
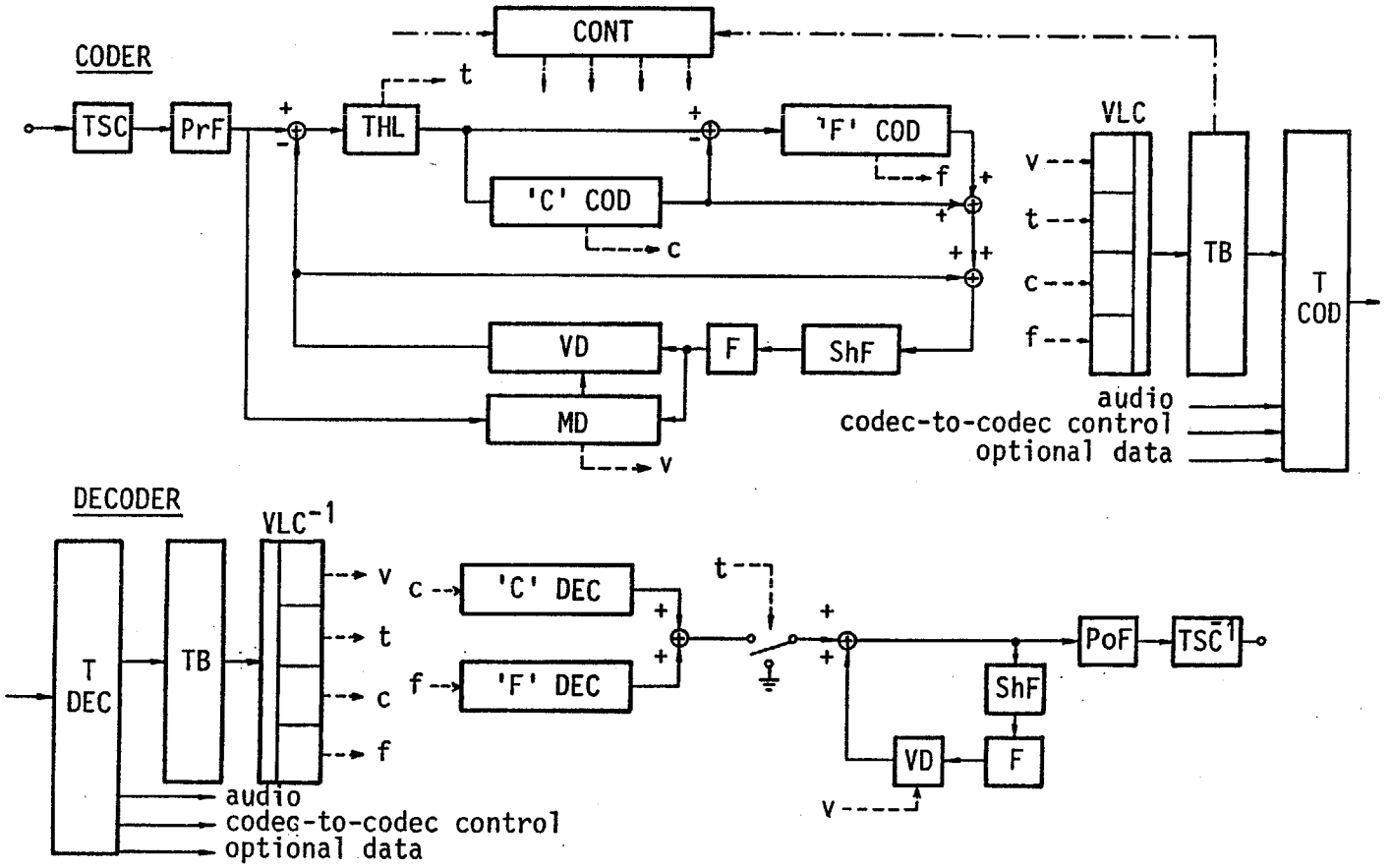


Figure 3 Temporal Pre-filter



Key

FUNCTIONAL BLOCK

TSC : Television Standards Converter to Intermediate Format
PrF : Pre-filter
MD : Motion Detector
VD : Variable Delay
F : Frame Memory
THL : Significant/insignificant Block Detector
'C' COD: C-component Encoder and Quantizer
'F' COD: F-component Encoder and Quantizer
ShF : Shaping Filter
VLC : Variable Length Coder
TB : Transmission Buffer
CONT : Coding Mode Controller
T COD : Transmission Coder

INFORMATION TO BE TRANSMITTED

v: Motion Vector
t: Flag for Significant/insignificant Block
c: C-component Quantizing Index
f: F-component Quantizing Index

Figure 1 A Generic Structure of 384 x n kbit/s CODEC