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

Document # 78
March 1986

SOURCE: NTT, KDD, NEC and FUJITSU

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

---

1. Introduction

At the last meeting in Ipswich, several candidate algorithms were presented to show that processed picture quality approximately reached the acceptable range for videoconferencing at 320 kbit/s. It was further agreed that the coding algorithm should be 'interframe prediction + Transform and/or VQ'.

This document proposes a generic structure of n x 384 kbit/s codec with Transform-based algorithm as a continuation of Documents #60 and #61, taking into account the recent study results on coding algorithm.

2. Coding Algorithm

We have been carrying out computer simulation of several variations of both DCT-based and VQ-based algorithms. The following items were proposed to be considered before finalizing the n x 384 kbit/s coding algorithm (see Section 4/#60).

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)

As for picture quality, DCT-based algorithm shows better performance at the present time, particularly in pictures containing quick and large motion such as Split-screen sequence. Performance of VQ-based algorithms has progressed rapidly, while it is not known whether it will finally surpass that of DCT-based algorithms or not.

As for hardware size and simplicity of compatibility related specifications, VQ-based algorithms will have advantage, because they include less amount of calculations in essential part of the algorithm.

As for adaptation to various input pictures and output rates, Transform-based algorithms have advantage because of more knowledge and experience.

Considering these together, we conclude to favor Transform-based algorithm for the n x 384 kbit/s codec with expectation of future microelectronics advancement to ease hardware related problems.

Annexes 1 and 2 show the latest computer simulation results for both DCT-based and VQ-based algorithms. Processed pictures will be demonstrated with video tapes at the meeting.

3. Codec Structure

3.1 Outline

Based on the fundamental configuration shown in Figure 1/#60 and Figure 1/#61, a more detailed version is proposed as Figure 1. Functional blocks without description are the same as in the previous documents.

3.2 Transform blocksize

According to the experience of computer simulation concerning DCT-based algorithms, we are in favor of 8 x 8 blocksize than 16 x 16 blocksize with the following reasons;

1) Coding efficiency

From the experience of carrying out coding simulation with both 8 x 8 and 16 x 16 blocksize, it is felt that 8 x 8 is equal to or slightly better than 16 x 16 in coding efficiency if we total the bits required for expressing significant blocks, overhead information and significant block ratio.

2) Picture quality

Blocksize of 8 x 8 gives better picture quality. In Transform-based coding, quantization distortion is dispersed through the whole block. Distortion is more objectionable with larger blocksize.

3) Hardware size

Smaller blocksize is obviously advantageous.

3.3 Adaptive quantization and coding

Blocks consisting of interframe prediction errors have a great variation, from those of still area with almost all elements being zero to those of changed scene with almost all elements being similar to original pels. Adaptive processing is essential in the algorithm of 'interframe prediction + transform'.

From our experiments, we propose the following adaptive scheme. Backgrounds of this proposal is presented in the companion document #79.

1) Each interframe prediction error block is classified according to its characteristics.

2) According to the above classification, the following two parameters are switched in a predetermined way for each block.

   a. A zone is decided, outside of which no transform coefficients are transmitted.
   b. A code set for entropy coding is switched from coefficient to coefficient inside the zone for each class.

3) In addition to the above, some coefficients are dropped which are redundant from human-observer characteristics.

3.4 F-coder

After evaluating the performance improvements and necessary hardware for F-coder, which was introduced in Document #60, we are of the opinion that it can be dealt with as an option. A detailed scheme will be proposed before the next meeting when we will have finished experiments concerning full resolution graphics for the agreed test sequence and operation for n greater than 1.

Study results and considerations leading to this conclusion are presented in the companion document #70.

3.5 Conditional motion-compensated frame interpolation

Conditional motion-compensated frame interpolation (CMI) improves the temporal resolution of decoded pictures, while it degrades transmission delay and necessitates additional hardware. We are of the opinion that efforts should first be directed toward improving coding efficiency to increase the number of frames to be coded and CMI should be introduced if there are not any other means. Consequently, it is dealt with as an option in Figure 1.

Details are discussed in the companion document #71.

3.6 Video information to be transmitted

As a summary, the video source coder having a generic structure mentioned above outputs the following three different kind of essential information, which are transmitted to the decoder through transmission channel.

a. Flag to identify whether a block to be processed is significant or not
b. Block classification flag for each significant block
c. Quantization index for each transform coefficient in the zone

The video source coder also outputs the following three different kind of optional information if corresponding options are adopted.

d. Motion vector for each block of motion compensation
e. Quantization index of F-coder
f. Quantization index of CMI-coder

3.7 Color difference signal processing

At the last meeting, whether color difference signals be processed with separate coder or with the same coder as that for luminance signal is recognized to require further study. It is also confirmed that TDM format is left open for hardware implementation, and that compatibility is ensured through the specification to the video multiplex.

When implementing the first generation codec and carrying out simulation study for the second generation codec, we have not found any serious problems in processing color difference signals using TDM format. Hence we are intending to adopt the same approach for the second generation codec hardware with such TDM format as shown in Figure 2. We request this would be considered in defining video multiplex coder.

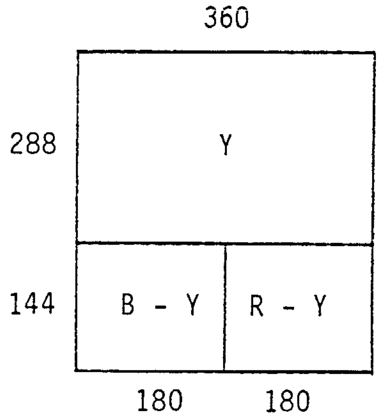


Figure 2 A TDM format for common intermediate format

3.8 Compatibility with m x 64 kbit/s codec

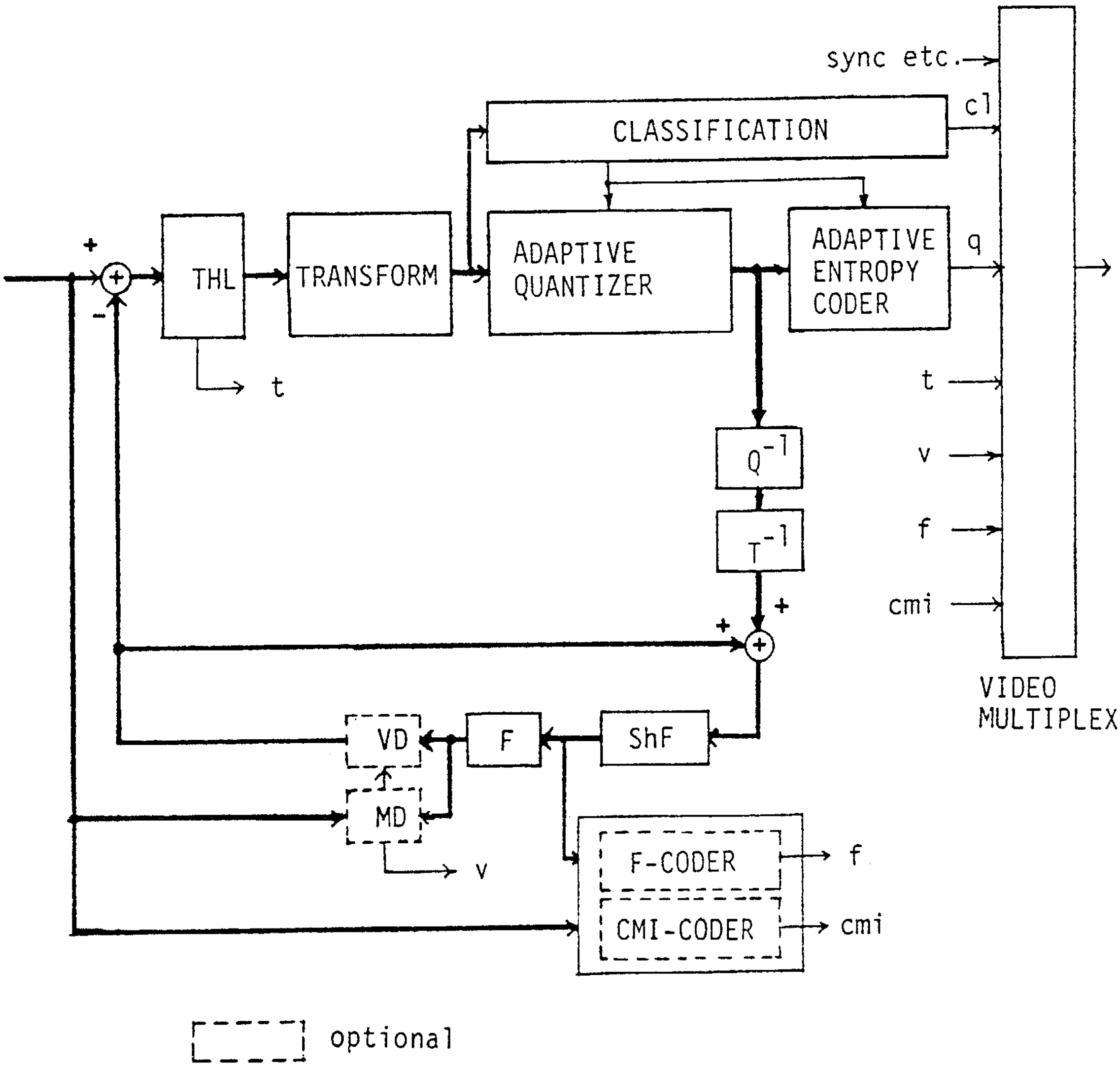
We propose that the standardization of n x 384 kbit/s codec should be given first priority and completed as soon as possible by dealing with the m x 64 kbit/s codec separately for the moment, since the service concept of the system using 64 kbit/s codec is not so clear now.

4. Conclusion

A generic structure of n x 384 kbit/s codec with Transform-based coding algorithm has been proposed. The detailed method of adaptive quantization and coding is to be decided. Next important work of the Specialists Group should be to specify the experimental hardware as soon as possible, for verifying effectiveness of the new coding algorithm and for checking compatibility among different design codecs. Optimization of the parameters should also be carried out with simulation and later with experimental hardware.

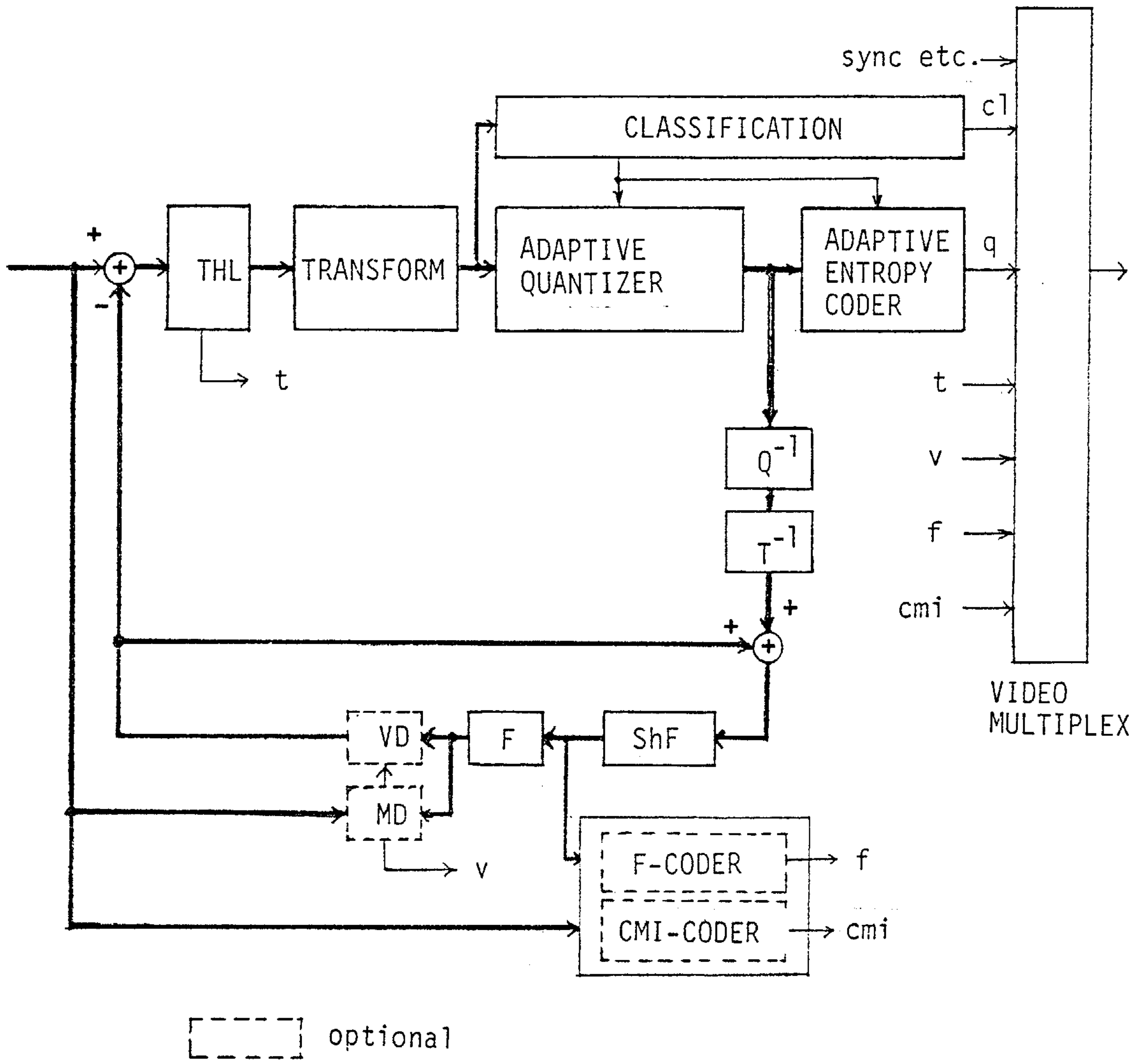
Annex 1

Annex 2



t : Flag for Significant/insignificant Block
cl : Classification Flag
q : Quantizing Index for Transform Coefficients
v : Motion Vector
f : F-coder Output
cmi: Conditional Motion-compensated Frame Interpolation Coder Output

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



t : Flag for Significant/insignificant Block
cl : Classification Flag
q : Quantizing Index for Transform Coefficients
v : Motion Vector
f : F-coder Output
cmi: Conditional Motion-compensated Frame Interpolation Coder Output

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

ANNEX / Doc. # 78

AN EXAMPLE OF SIMULATION RESULT OF DCT-BASED CODING ALGORITHM

1. Introduction

The simulation results demonstrated in the Tokyo meeting are obtained under the generic coding algorithm described in the document # 78 and # 79. This annex describes the specific coding algorithm and parameters used in the simulation.

2. Coding Algorithm

Motion compensated interframe difference signals are first segmented into blocks. Each block is tested whether it is significant or not. Only significant blocks are discrete cosine transformed. The blocksize is 8x8.

Figure A.1 shows the coding part for DCT coefficients. Each transformed block is classified into one of pre-determined classes. DCT coefficients are coded by using the class information to improve the coding efficiency and the resultant image quality.

In order to simplify the classification process, following procedure is adopted :

1) Detect significant coefficients.
   Binarization using threshold Th is performed to detect significant coefficients.
2) Find the smallest rectangle which surround the detected significant coefficients.
3) Select one of pre-determined patterns which corresponds to the above rectangle zone. These pattern is used to improve the coding efficiency.
4) Only coefficients inside the pattern are quantized. Coefficients outside the pattern are forced to zero and not coded.
5) Quantization results are Huffman coded. It is also planned to change quantization characteristics according to the class information in order to improve the overall performance.

Furthermore it may be useful to prepare several bit allocation tables for DCT coefficients and to select an appropriate table using the class information. However, this technique is not included in the simulation currently for the sake of simplicity.

3. Simulation Parameters

3.1. Pre-processing

The temporal filter and spatial filter are employed to reduce noise in the input signal. The necessity of the spatial filter is now under the investigation.

3.2. Detection of significant/insignificant block

A block of which the power of motion compensated prediction error is greater than threshold Th_p is judged to a significant block. The threshold Th_p is varied to control the rate of significant block in a coded frame.

3.3. Adaptive Quantization

The threshold to detect significant coefficients is varied to control the amount of generated coding information.

The number of classes is 16. The smallest pattern contains only the DC coefficient. On the other hand, the largest pattern contains almost all coefficients.

3.4. Frame Dropping

Frame dropping rate is normally 2:1. This rate is controlled depending on the buffer memory occupancy. For a scene change, higher frame dropping rate is employed to assign sufficient amount of bits to the changed scene.

3.5. Closed Loop Operation

Rate of significant blocks in a frame and buffer memory occupancy are used to control the amount of generated coding information. Parameters to be controlled are as follows :

- threshold to detect significant/insignificant blocks
- threshold to detect significant DCT coefficients
- quantization parameters
  dead zone and step size

3.6. Transmitted Information and Bit Assignment

The following information are transmitted :

- significant/insignificant block identification flag
  : Fixed word length coded. In the simulation insignificant blocks with v = 0 is distinguished from other blocks.
- motion vector
  : Variable word length coded. The code table is based on that of Part 3 codec.
- classification index
  : Fixed word length coded.
- scalar quantizing index
  : Fixed word length coded for DC coefficients in the scene change mode. Variable word length coded for DC coefficients in the normal mode and for AC coefficients in both the normal and scene change modes.

3.7. Post-processing

In the simulation a median filter with cross-shaped window (5 elememts) is employed as a spatial filter to reduce noise in the local decoded picture. This filter does not affect the performance of the coding loop.

4. Results of the Simulation

Test sequences are

A. Miss America

B. Checked Jacket

C. Split-Trevor.

Table A.1 shows an example of bit assignment to each transmitted information and signal to noise ratio of decoded signals (S_{pp} = 255).

Table A.1 Bit Assignment and S/N Ratio

Test Sequence	A. Miss America	B. Checked Jacket	C. Split	C. Trevor
Frame No.	61	40	29	99
S/N [dB]	38.39	38.60	35.26	37.02
Sig./Insig.	2571 (9.9%)	2392 (12.5%)	3067 (5.6%)	2650 (10.8%)
Motion Vector	4340 (16.8%)	1995 (10.4%)	10113 (18.4%)	4694 (19.1%)
Classification Index	1868 (7.2%)	1716 (9.0%)	3964 (7.2%)	2052 (8.4%)
Scalar Quant. Index	17099 (66.1%)	13047 (68.1%)	37831 (68.8%)	15120 (61.7%)
Total	25878 (100.0%)	19150 (100.0%)	54975 (100.0%)	24516 (100.0%)

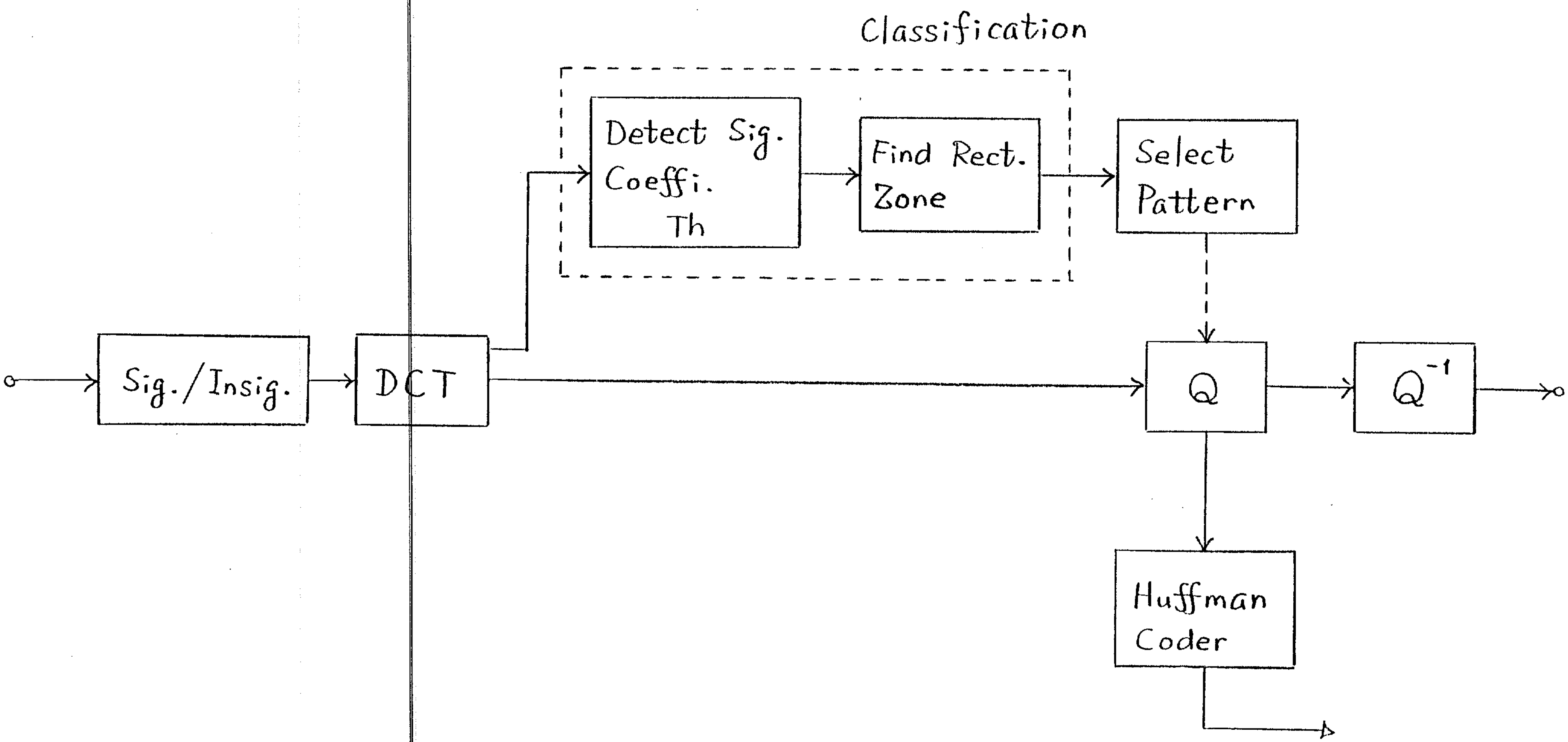


Figure A.1   Coding of DCT coefficients.

Annex 2 / Doc. #78

Simulation results of VQ-based coding: VCR

This paper describes simulation results of VQ-based codec. Coding parameters and mode control are the same as annex to document #62 of Ipswich meeting, main difference is that the number of mode control is changed from 3 to 5. Comparison between latest results and previous one at Ipswich meeting is shown in Table 1.

Table1. Simulation Results

Test sequence	Number of coding frames	SNR [ref. frame]
Miss America	68 / 150	39.6 [60th]
	63 / 150	39.5 [60th]
Checked Jacket	33 / 59	38.5 [40th]
	24 / 59	38.1 [40th]
Split Screen	11 / 60	35.0 [30th]
	11 / 60	32.2 [32nd]
TREVOR	24 / 89	37.3 [101st]
	22 / 89	36.8 [99th]

up : latest simulation results
down: previous simulation results at Ipswich