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

Doc no # 90
March 20, 1986

Source: SWEDEN

Title: SWEDISH HYBRID CODING ALGORITHM

In this document we present details concerning the latest simulations for video conferencing at 60 and 300 kbit/s using 8*8 DCT based hybrid coding with motion compensation.

Transform

The transform used is the classical Discrete Cosine Transform. The normalization constants make the transform orthonormal. 8*8 block size is used for both 60 and 300 kbit/s. The dynamic range of the dc component is +/- 2048.

$$F(u,v) = \sum_{x=1}^{N} \sum_{y=1}^{M} C(u)\,C(v)\, f(x,y) \cos \frac{\pi u (x-\frac{1}{2})}{N} \cos \frac{\pi v (y-\frac{1}{2})}{M}$$

$$f(x,y) = \sum_{u=1}^{N} \sum_{v=1}^{M} C(u)\,C(v)\, F(u,v) \cos \frac{\pi u (x-\frac{1}{2})}{N} \cos \frac{\pi v (y-\frac{1}{2})}{M}$$

where (x,y) are pel domain, and (u,v) are transform domain coordinates. C(0) = 0.707 and C(w) = 1 for w > 0. M and N are both 8.

Motion compensation

A set of 32 different motion vectors is used,
requiring 5 bits of side information,
(plus 1 bit indicating if block is moving/non-moving,
see also comment below under "Variable length coding").
The following motion vectors are used.

	-8	-4	-2	-1	0	1	2	4	8
-8	X				X				X
-4		X			X			X	
-2			X		X		X		
-1				X	X	X			
0	X	X	X	X	Z	X	X	X	X
1				X	X	X			
2			X		X		X		
4		X			X			X	
8	X				X				X

For high frequencies the above vectors give poor prediction. Therefore a blurring filter is used on previously reconstructed picture before prediction. This filter is only used on moving blocks. The filter is a linear 5 tap FIR-filter with the following coefficients.

	1/8	
1/8	1/2	1/8
	1/8	

The motion estimation is done by full search among the above set. The criterion is mean magnitude. In order to use motion vectors only when they give improvement, a prefference is given to the zero vector:

"magnitude zero vector" < 1.25 * "magnitude best vector"

=> chose zero vector for prediction

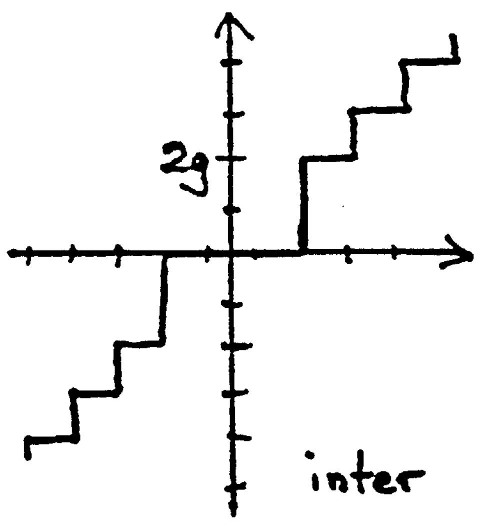
Motion estimation is done using previously coded picture.
Only the luminance signal is used.
The same vectors are used (properly scaled) in the corresponding areas of the chrominance signal.
This gives improved prediction for the chrominance without requiring extra side information.

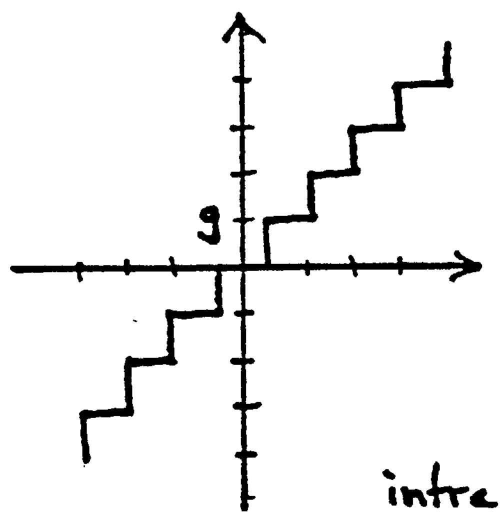
Numerical results

	FD	MCFD	no of bits/pictu for motion vecto
SPLIT/TREVOR 10 Hz 300 kbit/s	19.1	9.6	3200
MISS AMERICA 7.5 Hz 60 kbit/s	11.5	5.3	2200

Quantization

All transform components in a block are quantized with the same quantizer, (for buffer control, see below). The quantizer is uniform with a large number of levels, coverring a large portion of the dynamic range of the transform. In inter-frame coding the +/- g level is removed, giving a threshold for small levels. This threshold has two advantages. It works as a hysteresis, making the coder less sensitive to camera noise. The bit rate also becomes considerably lower with this threshold.
The threshold is not used in intra-frame coding.





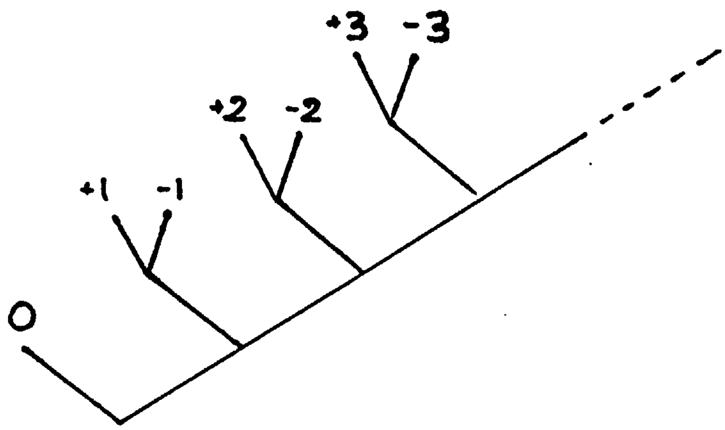
Variable length coding.

All blocks are transformed and quantized, i.e. no detection of "changed blocks" is done. However, a large percentage of the blocks only contain zeroes after quantization. These blocks are efficiently coded with one bit, saying "only zeroes", which is equivalent to "uncoded". The blocks which have at least one non-zero component after quantization are coded with the following sceme.

Note: The block attributes "moving/non-moving" and "coded/uncoded" normally require one bit each per block. However, this binary information has been further compressed with a variable length coding algorithm called "arithmetic coding", described in the appendix of doc no # "VIDEO MULTIPLEX FOR n*384 KBIT/S" by UK, Sweden and France.

(1) Components
==================

In inter-frame coding, every quantized transform component is variable length coded with the same tree code, see figure. In intra-frame coding the same holds, except for the DC component which is fix length coded with the necessary number of bits to cover full dynamic range of the quantizer. Note that run-length coding of zeroes is not used. The zeroes are coded with one bit each.



(2) End of block
====================

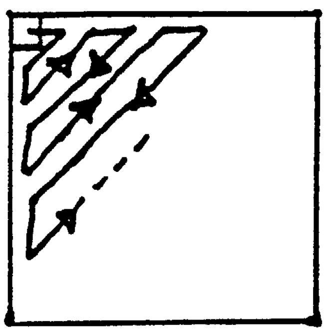
It suffices to indicate to the decoder how many non-zeroes there are in a block. This is done with a special variable length word before the component bits. The number of non-zeroes in a block is maximized. In intra-frame mode the maximum is 8 non-zeroes, and in inter-frame mode the maximum is 31. No block with 31 non-zeroes has been found in inter-frame mode.

Numerical Results
=================

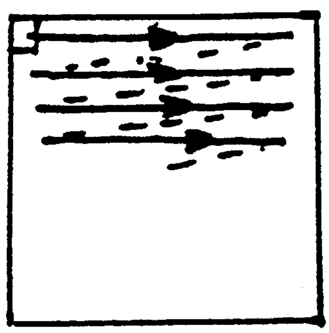
SPLIT/TREVOR	300 kbit/s	3.5 bits/block for no of non-zeroes
MISS AMERICA	60 kbit/s	2.5 bits/block for no of non-zeroes

(3) Scanning

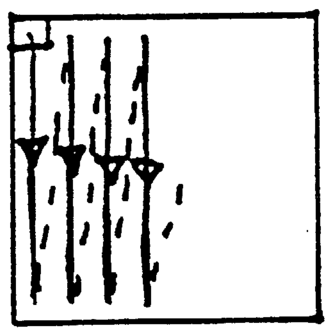
It is desirable to hit all non-zeroes with as few zeroes as possible between them (after the last non-zero the scanning is stopped). An efficient way of scanning is the so called zig-zag scanning.
In some blocks it might be prefferrable to scan in another way, due to he main frequency content of the block. Experience has shown that for 8*8 blocks it suffices with the three classes shown below.
Intruducing more classes gives more efficient scanning, but also costs more side information which was not payed back.



"zig-zag"



"H"



"V"

Results with SPLIT/TREVOR at 300 kbit/s

The introduction of "H" and "V" classes saved about 2000 bits/picture, compared to using only "zig-zag".

Buffer control

In all simulation a buffer size corresponding to one picture has been used (i.e. 30 kbit for 10 Hz/300 kbit/s, 8 kbit for 7.5 Hz/60 kbit/s etc.).

The buffer content controls the quantizer stepsize. Eight different stepsizes are used, namely
g = (4, 8, 12, 16, 20, 24, 28, 32).
Quantizer choise is done for each "line of blocks". A "line of blocks" (LOB) is defined to consist of two lines of Y-blocks, one line of U-blocks and one line of V-blocks. Totally, an LOB consists of 134 blocks representing 16 lines of colored picture. Side information for quantizer choise is 3*18=54 bits/picture.
To cope with variable activity in different parts of the picture the frame difference is also taken into account for the quantizer choise, see the non-optimized table below. Scene cuts are detected with a threshold on frame difference. If "scene cut" is detected, the entire picture is coded "intra" with double number of bits, i.e. an extra frame drop. Otherwise it is always coded "inter" (although we believe "intra" should be used also for frame refresh).
This control structure requires knowledge of frame difference before coding can start. The frame difference is calculated when a new picture is read into the memory.

frame difference																	
2.	4	4	4	4	4	4	4	4	4	8	12	16	20	24	28	32	
4.	4	4	4	4	4	4	4	8	8	8	12	16	20	24	28	32	
6.	4	4	4	4	4	4	8	8	8	12	12	16	20	24	28	32	
8.	4	4	8	8	8	8	8	12	12	12	12	16	20	24	28	32	
10.	4	8	8	8	8	8	12	12	12	12	16	16	20	24	28	32	
12.	4	8	8	8	12	12	12	12	12	16	16	16	20	24	28	32	
14.	4	8	12	12	12	12	12	16	16	16	16	16	20	24	28	32	(inter)
16.	4	8	12	12	12	12	16	16	16	16	16	20	20	24	28	32	
18.	4	8	12	12	16	16	16	16	16	16	20	20	20	24	28	32	
20.	4	8	12	16	16	16	16	16	16	20	20	20	24	24	28	32	
22.	4	8	12	16	16	16	16	20	20	20	20	24	24	28	28	32	
24.	4	8	12	16	16	16	20	20	20	20	24	24	28	28	32	32	
28.	4	8	12	16	16	20	20	20	20	24	24	28	28	32	32	32	
32.	4	8	12	16	20	20	20	20	24	24	28	28	32	32	32	32	
34.	4	8	12	16	20	24	24	24	24	28	28	32	32	32	32	32	
	20	20	20	20	24	24	24	24	24	28	28	28	28	32	32	32	(intra)
	0%				25%				50%				75%			100%	buffer content