

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.

(2)

# 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	X	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,

$$\begin{array}{ccccccc} & & & & 1/8 & & \\ & & & & & & \\ & & 1/8 & & 1/2 & & 1/8 \\ & & & & & & \\ & & & & 1/8 & & \end{array}$$

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 preference is given to the zero vector:

$$\begin{array}{l} \text{"magnitude zero vector"} < 1.25 * \text{"magnitude best vector"} \\ \Rightarrow \text{chose zero vector for prediction} \end{array}$$

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.

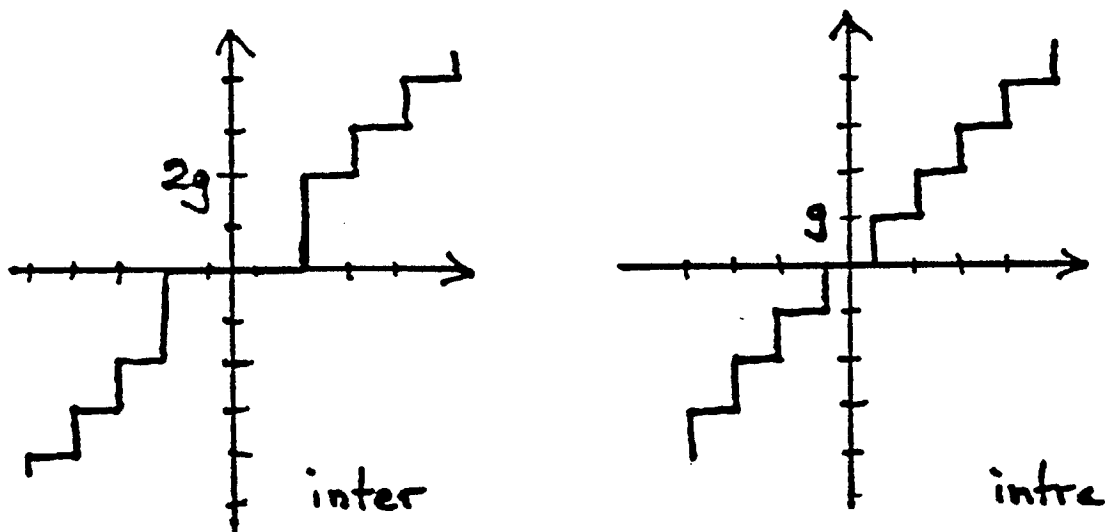
Numerical results =====			FD	MCFD	no of bits/pictu for motion vecto
SPLIT/TREVOR	10 Hz	300 kbit/s	19.1	9.6	3200
MISS AMERICA	7.5 Hz	60 kbit/s	11.5	5.3	2200

(3)

### 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, covering a large portion of the dynamic range of the transform. In inter-frame coding the  $\pm g$  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 scheme.

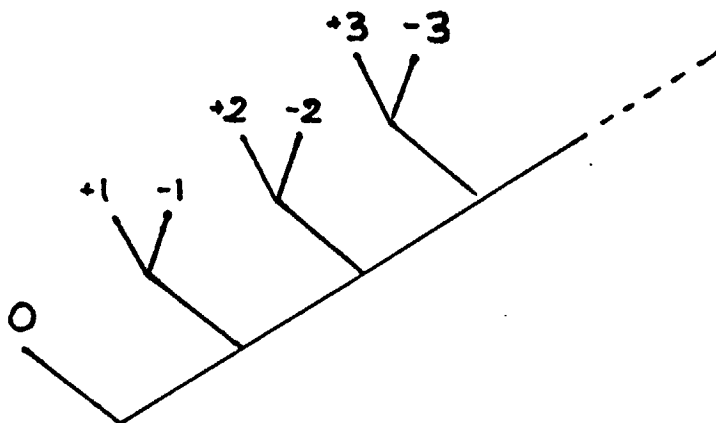
Note: The block attributes "moving/non-moving" and "coded/uncoded" normally require one bit each per block. However, this binary information has been further compressed with a variable length coding algorithm called "arithmetic coding", described in the appendix of doc no # "VIDEO MULTIPLEX FOR n\*384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

(4)

(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

11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 10

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

00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

SPLIT/TREVOR 300 kbit/s

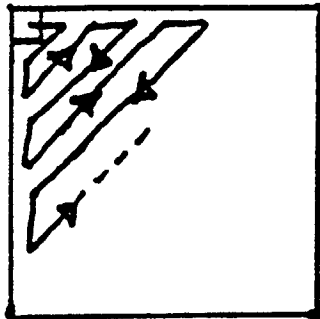
3,5 bits/block for no of non-zeroes

MISS AMERICA 60 kbit/s

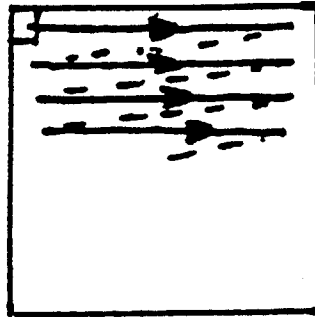
2,5 bits/block for no of non-zeroes

(3) Scanning  
=====

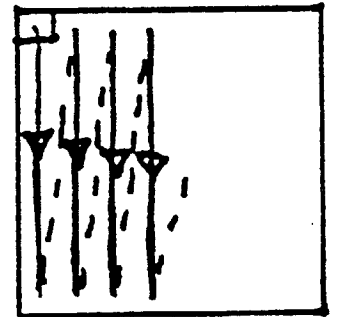
It is desirable to hit all non-zeroes with as few zeroes as possible between them (after the last non-zero the scanning is stopped). An efficient way of scanning is the so called zig-zag scanning. In some blocks it might be prefferrable to scan in another way, due to he main frequency content of the 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 choice 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 choice is  $3 \times 18 = 54$  bits/picture. To cope with variable activity in different parts of the picture the frame difference is also taken into account for the quantizer choice, 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	16	20	20	24	28	32
18,	4	8	12	12	16	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	
<hr/>																	
	20	20	20	20	24	24	24	24	24	28	28	28	28	32	32	32	(intra)
0%	25%				50%				75%				100%				buffer content