

INTERNATIONAL ORGANISATION FOR STANDARDIZATION
ORGANISATION INTERNATIONALE DE NORMALISATION

ISO-IEC/JTC1/SG2/WG8

CODED REPRESENTATION OF PICTURE AND AUDIO INFORMATION

ISO-IEC/JTC1/SG2/WG8
MPEG 89/
December 1989

ISO/IEC JTC1/SC2/WG8
MPEG 89/248

Source: Dietmar Hepper
Leader of MPEG-Video Development Group I
Title: Coordination of Further Studies on Promising Algorithm
Components towards a Group I Reference Model
Purpose: Information for MPEG Video Development Group I members
Status: Draft #2

Coordination of Further Studies on Promising Algorithm
Components towards a Group I Reference Model

General

The list of promising algorithm components (see MPEG Video Group I document no. MPEG 89/..., Nov. 23, 1989) has been reviewed: action points have been grouped and numbers modified correspondingly for better clearness. In addition numbers reflecting priorities have been assigned and items that need to be standardized have been marked. Person names, rather than company names, have been assigned to the action points according to the information received from Group I members. The revised list is given below.

At the time being, Group I members are using more or less similar, however, different coding schemes for algorithm development and for comparison of algorithm components. By the January meeting in Eindhoven, agreement on algorithm components that should form a Reference Model, for proceeding towards an MPEG video coding standard, is expected within Group I and within MPEG Video as a whole.

Work on a Reference Model has been started in the Japanese branch of Group I and within Group II (see below).

Promising Algorithm Components and Task Assignment

Here is the revised list of promising algorithm components that need further study:

1. Coding structure (intra/predictive/interpolative/extrapolative coding):
 - a) group-of-frame structure (period of intra-coded frames/frequency predicted/interpolated frames (adaptive?) ...)
 - b) criterion for interpolation error transmission
2. Coding of intra frames (CCITT/JPEG, quantization, etc.)
3. Motion estimation for motion-compensated prediction and motion compensating interpolation:
 - a) block motion estimation scheme (hierarchical/non-hierarchical ...)
 - b) basis: source/reconstructed images
 - c) motion vector accuracy: fractional/integer pel accuracy; for prediction, interpolation
 - d) global motion estimation
 - e) motion vector range
4. Motion compensation (global and local):
 - global MC (spatial interpolation for fractional pel MC)
 - local "on" global MC vs. local/global MC selected on a (macro)block basis
5. Block sizes for motion vectors (for prediction/interpolation), mode selection, etc.
6. Filter in the loop:
 - a) filter after the frame memory (coefficients, criterion, pan!)
 - b) filter before the frame memory for blockiness reduction
7. AC prediction
8. Adaptive quantization/weighting of DCT coefficients
9. Block types
10. VLC:
 - a) coding of coefficients (CCITT/JPEG/others)
 - b) arithmetic coding
 - c) motion vector coding
 - d) block type coding
11. Coder control strategy:
 - a) bit assignment to the intra-coded/predicted/interpolated/extrapolated frames
 - b) control of quantizer step size
 - c) control of maximum number of transmitted coefficients
 - d) basis of control (group of blocks/macro block/...)
12. Pre- and postprocessing:
 - a) format conversion CCIR <-> CIFxy (filter coefficients, fixed/motion-adaptive/motion-compensated)
 - b) noise reduction
 - c) blockiness reduction
13. Chrominance resolution/source format(s)
14. Chrominance coding
15. Data structure (see 10.)

The current status of task assignment to Group I members is as follows (used abbreviations of person names are explained one page later):

new (old number)	Pr io	St nd	action point	under investigation by
1. a (3) b (5)	++ +	S	Coding structure: group-of-frame structure criter. for int. err. Xmission	TKum, JvdM, APur, JYun CGon, (BHam,) JvdM, APur, JYun
2. (7)	+	S	Intraframe coding	HWat
3. (2) a b c (4) d (1) e	 + o + ++ o	 S S	Motion estimation: block motion est. scheme basis: source/reconstructed motion vector accuracy global motion est. scheme motion vector range	CGon, BHam, YHon, TKog, TKum CGon, BHam, YHon, TKog, TKum, APur YHon, JvdM DHep, TKog, TKum
4. (1, 20)	 + +	 S S	Motion compensation global MC local "on"/parallel to global	TKog DHep (, RtHo) DHep (, RtHo)
5. (16)	+	S	Block sizes (vectors, mode ...)	YHon, APur
6. a (12) b (9)	 o +	 S S	Filter in the loop: filter after the frame memory filter before the frame memory	TMat NTex, CGon
7. (10)	+	S	AC prediction	NTex, CGon DLGa????
8. (8)	+	S	Quantization/weighting	HWat, PTou, RtHo
9. (18)	o	S	Block types	TKog, JYon, JvdM
10. a (13) b (14) c (15) d	 + o + +	 S S S S	VLC: coding of coefficients arithmetic coding motion vector coding block type coding	BHam, GMor, APur, HWat CGon APur, JYon
11. (6) a b c d	++ ++ + o o	 S	Coder control strategy: bit assignment to frames control of quantizer step size control of max. no. of coeff. basis of control	DHep, TKog DHep, JvdM, APur DHep, APur DHep
12. a b (17) c (11)	 + o o	 	Pre- and postprocessing (Group III!): format conversion noise reduction blockiness reduction	DHep, GMor YKat, YYam YKat
13. (19)	+	S	Chrom. resol./source format(s)	YKat, JvdM
14.	+	S	Chrominance coding	JvdM
15.	+	S	Data structure (see 10.)	JvdM

Prio = Priority, Stnd = need to be standardized
for abbreviations of names see next page

Following abbreviations of names of Group I members are used, each representing also his/her colleagues:

Adolph, Dirk	DAdo
Butera, Bill	BBut
Gonzales, Cesar	CGon
Guglielmo, Mario	MGug
Hammer, Bernard	BHam
Puri, Atul	APur
Hepper, Dietmar	DHep
Hontama, Yasukazu	YHon
Katayama, Yasuo	YKat
Kogure, Takuyo	TKog
Kummerow, Thomas	TKum
Le Gall, Didier	DLGa
Matsuda, Kiichi	KMat
Matsuyama, Takeshi	TKog
Meer, Jan van der	JvdM
Morris, John	JMor
Morrison, Geoffrey	GMor
Saint Girons, Régis	RSGi
Simon, Allen	ASim
ter Horst, Roel	RtHo
Texier, Nicole	NTex
Tourtier, Philippe	PTou
Watanabe, Hiroshi	HWat
Yamazaki, Yasuhiro	YYam
Yonemitsu, Jun	JYon

Coding Structure

As regards coding structure, a general scheme for MPEG Video could be described as follows using two parameters N and M with reference to the coding method used (see Fig. 1):

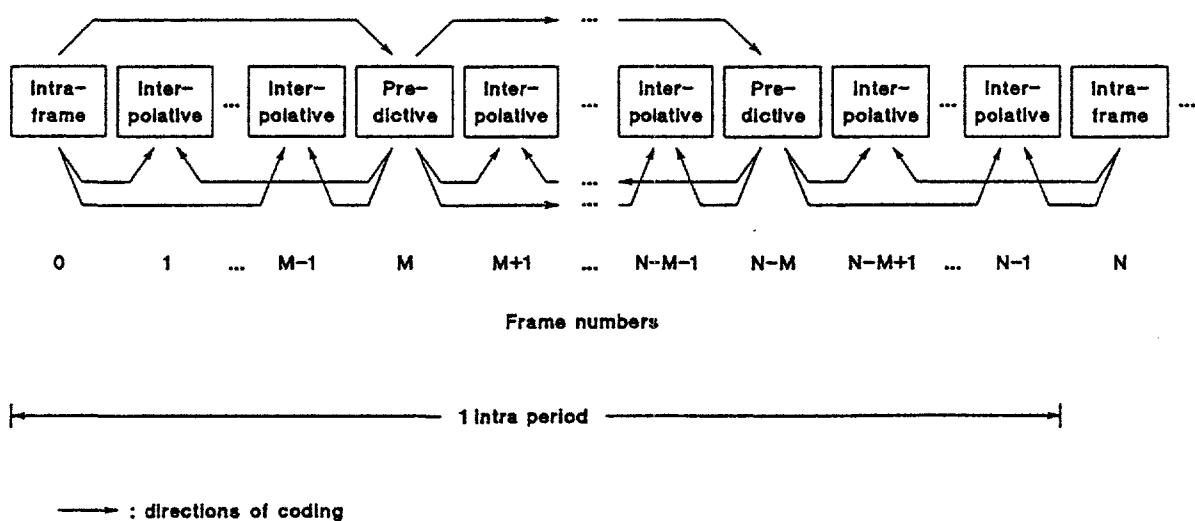


Fig. 1 General coding structure

N = distance of intraframe coded frames

M = distance of unidirectional, i. e. predictively coded frames

- intraframe coding:
N = distance of frames coded in intra mode
e. g.: N = 15, 12, 10, or 8
- predictive coding:
M = distance of frames coded only unidirectional, i. e. predictively
e. g.: M = 2
- interpolative coding:
remaining frames are coded also bidirectional, i. e. using interpolation

E. g., with the assumed restriction of the maximum random access time being below 0.5 sec, possible parameters of N and M are:

	N	M
50 Hz	12	2... 6
60 Hz	14 or 15	2... 7

The basic understanding of the focus of work within Group I is M = 2. The Japanese branch of Group I focusses on N = 15, M = 3 for a Reference Model whereas Group II is currently concentrating on N = 6, M = ∞. Fig. 2 shows the coding structure for these examples of certain values of parameters N and M.

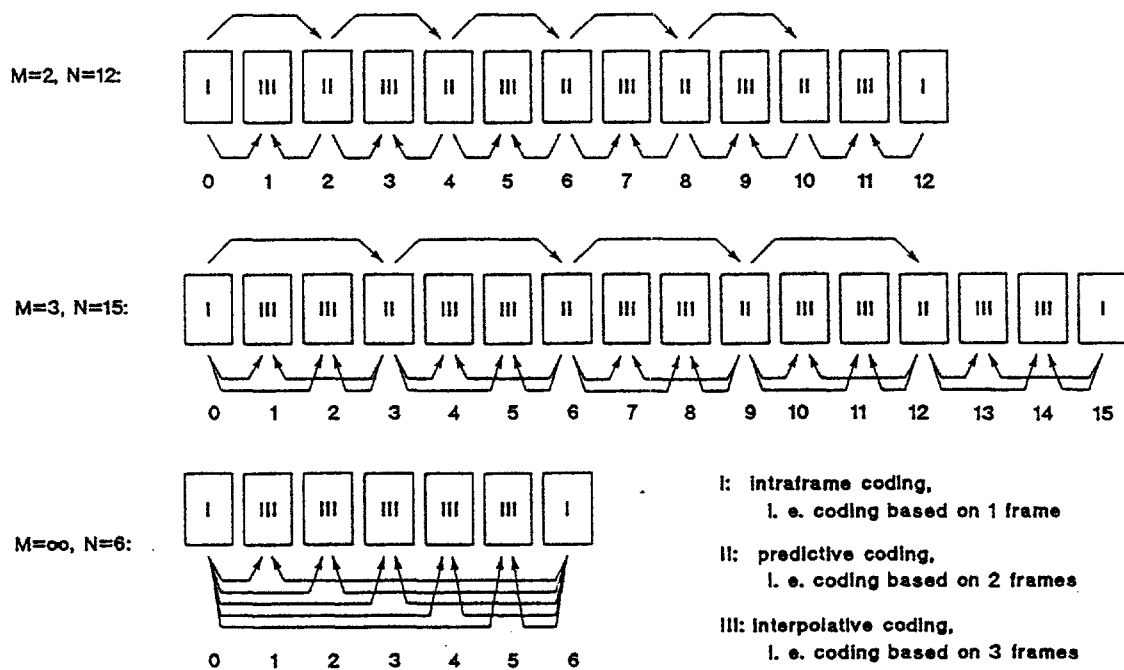


Fig. 2 Examples of coding structure, specified by parameters N and M

Information on Japanese Group I Work

Mr. Kogure has installed a Japanese Group I with coordinated activities on the 20 action points listed in the summary of promising algorithm components of November 23. A copy of the fax received from Mr. Kogure is attached to this one. A Reference Model has been worked out - and information on it distributed - to serve as a basis for their further algorithm components comparisons. It is strongly based on CCITT RM8/H.261:

- input format currently 30 Hz, Y: 352*240, C: 176*120
- vertical-temporal non-linear filtering for format conversion
- group-of-frames structure:
 - 2 GOFs/sec, 1 GOF = 15 frames
 - every 15th frame intra-coded (1, 16, ...) [N = 15]
 - 4 frames in between (4, 7, 10, 13) coded with unidirectional, motion-compensated prediction (4 from 1, 7 from 4, etc.) [M = 3]
 - remaining 10 frames coded bidirectional with motion-compensated prediction/weighted interpolation (2 and 3 from 1 and 4, etc.)
- group-of-block and macroblock structure as in RM8
- motion compensation:
 - 1 vector per macroblock
 - motion vector estimation for luminance by full search using previous, current and next source images
 - maximum vector range, horizontally and vertically:
 - for interpolated frames: ± 7 pels
 - for predicted frames: ± 21 pels (estimated within a range of ± 7 pels starting from twice the average of the vectors for interpolated frames in between)
 - chrominance vectors derived by halving the components of the Y ones (INT operation)
 - MC/NoMC decision as in RM8
 - loop filter:
 - coefficients as in Bellcore and GCT proposals,
 - on/off control not precisely defined, local pan detection used as a default: MB filtered if motion vector non-zero and MB does not move together with neighbouring MBs (pan criterion: vector comparison with 8 neighbouring MBs, if the number of identical vectors is greater or equal to 4 [recommended value] the MB is regarded as panning)
 - application only to frames with unidirectional, i. e. predictive coding
 - Global motion compensation is under study but not yet included in the reference model.
- prediction mode decision (previous/interpolate/next) based on minimum error
- intra mode decision: as in RM8
- quantization:
 - weighting of coefficients (no weighting used at the moment)
 - uniform quantization with dead-zone as in RM8
 - quantizer step size control under study (selected according to the buffer fullness)
 - for intra-coded frames: unique step size q; furthermore for each Y block within a macroblock: 4 different modes application, adapted according to an activity measure:

$q/4$ applied if $Th_0 < Y_{Act} \leq Th_1$ and $Y(0,0) > Th_{DC}$
 $NINT(q/3)$ applied if $Th_1 < Y_{Act} \leq Th_2$ and $Y(0,0) > Th_{DC}$
 $q/2$ applied if $Th_2 < Y_{Act} \leq Th_3$ and $Y(0,0) > Th_{DC}$
 q applied otherwise

with:

$$Y_{Act} = \sum_{u=0}^7 \sum_{v=0}^7 Y(u,v)^2 - Y(0,0)^2$$

$Y(u,v)$: DCT coefficients

Th_{DC} : threshold for DC coefficients

Th_n : n-th threshold for Y_{Act} ($n=0, 1, 2, 3$)

- DC quantization in intra mode: uniform with step size 8:

$$Q(DC) = INT((DC-4)/8)+1,$$

transmitted with 8-bit FLC

- chrominance block: Y step size multiplied with a certain factor, e. g. 0.5

- VLC:

- 2-dimensional VWL coding of AC coefficients as in CCITT H.261
- VLC coding of motion vector differences
- Coded Block Pattern as in RM8
- Macroblock addresssing as in RM8, modification of the meaning of "fixed MBs": all coefficients zero, zero motion vector data, side information (MB type, loop filter control information, quantizer step size, etc.) identical to the previous MB

- data rate control: under study

- On the list of action points is an adaptive blockiness removal filter which is applied to pixels at the block boundaries:

$$X_0 \left| \begin{array}{cc} X_1 & X_2 \\ \text{block boundary} & \end{array} \right. \quad \text{values before filtering}$$

$$F = (3 X_0 + X_1 + 2 X_2)/6 \quad \text{or} \quad (2 X_0 + X_1 + X_2)/4$$

$$X_{1out} = X_1 + r d = (1-r) X_1 + r F$$

with $r = a / (a + |d|)$

$$d = F - X_1$$

$$a = \text{stepsize}/8.0/0.40$$

First filtering in x direction is carried out, the output is then filtered in y direction.

Information on Group II Work

Within Group II Bernard Szabo has undertaken to define a voluntary Reference Model. It is based on:

- pre- and postprocessing:
 - source format 525/30 Hz until January,
 - coded format: Y: 352*240, C: 176*120
 - preprocessing: field dropping, vertical and horizontal decimation
 - postprocessing: vertical and horizontal interpolation
- group-of-frame structure:
 - every 6th frame intra-coded [$N = 6$], intermediate frames bidirectional with four modes (intra (128), forward MC, backward MC, weighted average MC, i. e. interpolation) [$M = \infty$]

- mode selection:
 - on an macroblock basis
 - minimum square energy of difference to input used as selection criterion; when the MSE for zero motion is less than 4 per pixel (1024 over a macroblock) the motion vector is set to (0,0)
 - the importance of DC mismatch is lessened when deciding whether to use intra mode or not: for intra mode error measure following MSE is used:
$$\text{MSE}' = \text{MSE} - (\text{DC}-12) * (\text{DC}-128)$$
MSE measures of other 3 possible modes are retained
- motion compensation:
 - no global motion compensation
 - integer motion vector (MV) estimation on a 16*16, i. e. macroblock basis; estimates MV01 and MV65 serve as a basis for successive estimation of MV02, MV03, MV04, MV05 and MV64, MV63, MV62, MV61, respectively: each MV0(i) is estimated within a range of ± 7 pels around the current accumulated vector (MV01+...+MV0(i-1)); similarly the backward vectors MV6(i) are derived
 - fractional pel motion compensation for interpolation mode, vectors MV06 are divided according to the distance of the current frame from frames 0 and 6
 - all-pass 3*3 FIR filter for motion-compensated prediction (no filter at the moment)
- 8*8 DCT (after shift of values about zero [-128...+127])
- quantization:
 - weighting matrices
 - scaling factor f for rate control; for DCs the lesser of ACs' f and 10 is used
- zig-zag scanning
- VLC:
 - VLC for block types
 - JPEG VLC for coefficients:
 - differential coding of the quantized DC values of intra-coded frames
 - codes for non-zero ACs: zero-runlength/amplitude word length (2-dimensional VLC) plus amplitude (2, variable-length integers)
 - fixed/variable length motion vector difference coding depending on coding, i. e. MC mode; maximum of one forward vector and one backward vector per macroblock
- data rate control:
 - causal bit rate allocation algorithm, i. e. forward, no iterations
 - only one f-factor (scaling factor for basic quantization matrix values) used per frame
 - variable low-pass filtering (α -factor) as rate control tool
 - rate control performed on a GOF basis
 - initial values for α and f:

sequence	α	f
Table Tennis	2/16	65
Flower Garden	6/16	100
Diva	0	40

- control of α and f :
 next α = current α + $\alpha' * B$
 next f = round(current $f * 2^B$)
where $\alpha' = 8/16$
 B = buffered bits/128 kbits

DTB / Hannover
CHD-dgh