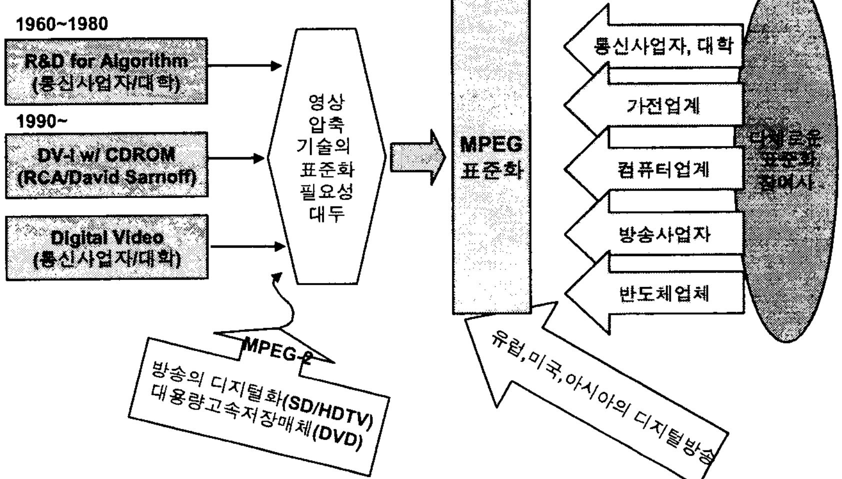
MPEG-2 & MPEG-4 Video Coding Techniques

November 23, 2001

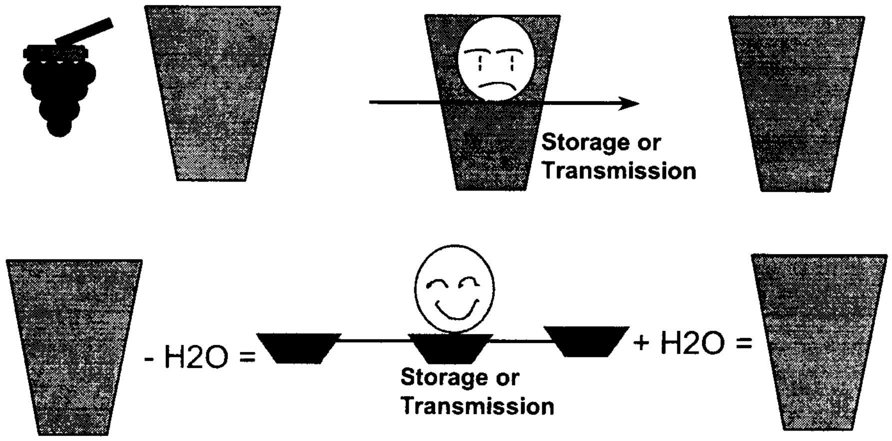
Byeungwoo Jeon
Digital Media Lab.
Sung Kyun Kwan University
bjeon@yurim.skku.ac.kr, (031)290-7144

MPEG-1,2 표준화의 배경

- MPEG(Moving Pictures Coding Experts Group)



Basic Idea of Compression



Point of Video compression

- How to reduce data with minimum signal distortion ???

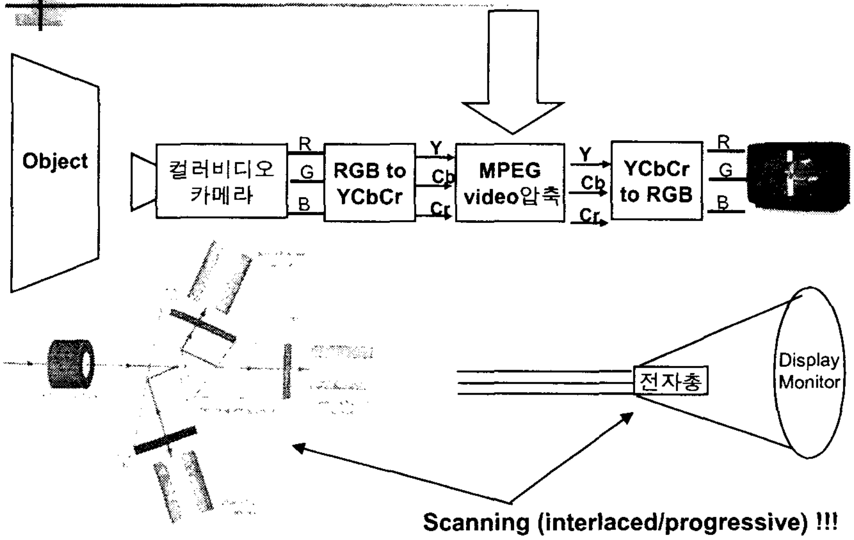
Technically speaking....

- ***Information to utilize in compression :***
  - ***A. Statistical Redundancy : redundant information***
    - ***- Variable - length coding***
  - ***B. Spatial & Temporal Redundancy : predictable information***
    - ***- DPCM, Motion Compensation(MC)***
  - ***C. Human Visual System : information which human cannot honor***
    (Ex. Human eye is less sensitive to high frequencies)
    - ***- DCT and Weighted Quantization***
    - ***- Spectral (colors) : color space conversion + sub-sampling***
- ***A, B : decoder can re-generate it***
- ***C : decoder does not re-generate it (human will not honor if it does)***

Keys in Digital Video Compression

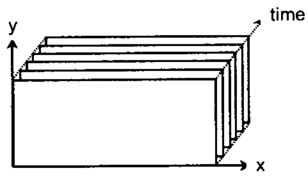
- *Scanning (progressive/interlaced, field/frame)*
- *Sampling (analog -> digital), color sub-sampling*
- *Quantization (Q/IQ)*
- *Transform (DCT/IDCT)*
- *Motion estimation/ Compensation (ME/MC)*
- *Variable-length Coding (VLC/VLD) - Huffman Coding*
- *Rate control*

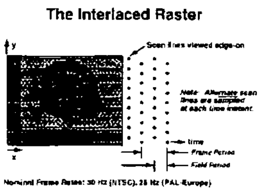
What we will Cover

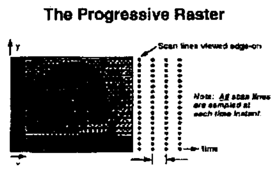


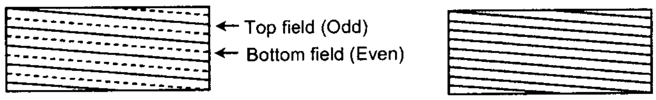
Scanning of Image Sequence

- Spatial (Horizontal/Vertical) & Temporal Scanning
  - ***- progressive/interlaced***
  - ***- field / frame picture, field / frame rate***
  - ***- top / bottom field***









Interlaced Scan

Progressive Scan

Current analog TV (interlaced)

NTSC	**720 X 480 @ 60 field/sec**
PAL/SECAM	**720 X 576 @ 50 field/sec**

Progressive :
Computer
Film (24 frames/sec)

Lecture Note by B. Jeon(Nov.23,2001)

7

Color Space Conversion:RGB->YCbCr

- **CIE표준 컬러 좌표계 : Y, R-Y, B-Y**
  - **Y = 0.299R + 0.587G + 0.114B**
  - **B-Y = -0.299R - 0.587G + 0.886B** **(-0.886 ~ +0.886 : 1.772)**
  - **R-Y = 0.701R - 0.587G - 0.114B** **(-0.701 ~ +0.701 : 1.402)**

- **Color Difference Signal Scaling**
  - **Cb = 0.564 (B-Y) + 128**
  - **Cr = 0.713 (R-Y) + 128**

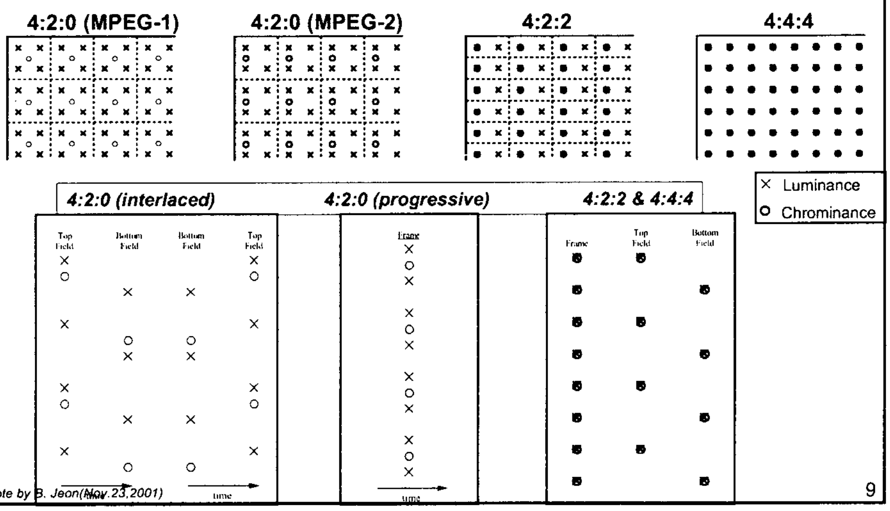
- **Signal Levels :**
  - **Y - 220 Level(16 ~ 235)**
  - **Color - 225 Level (16 ~ 240)**

Lecture Note by B. Jeon(Nov.23,2001)

8

Color Space Conversion & Sub-Sampling

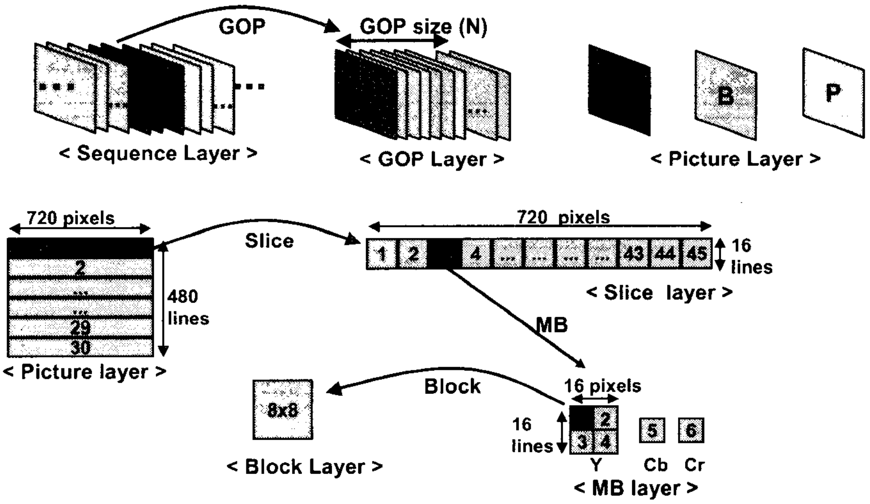
- RGB -> YCbCr: ***YCbCr is less correlated, Cb & Cr has smaller bandwidth.***
- ***Color Sub-Sampling***:



Lecture Note by B. Jeon(Nov.23,2001)

Layered Structure of MPEG-2

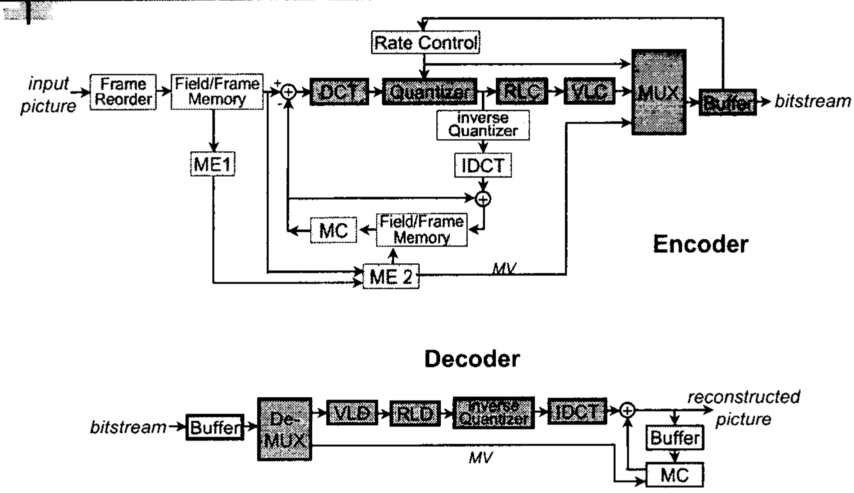
- Input Example : ITU-R 601 720(HT) X 480 (VT) - MP@ML



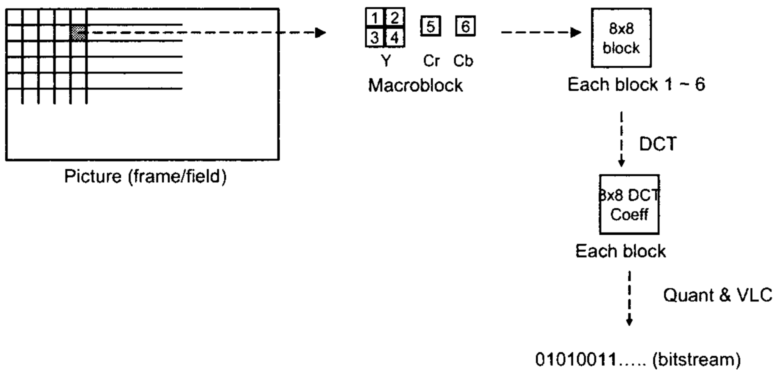
Lecture Note by B. Jeon(Nov.23,2001)

Concept of MPEG Intra Coding

Intra Coding of MPEG-2



Procedure of Intra Coding



What is & why DCT transform?

- **Given 8x8 block is represented as,**

= F(0,0) x A(0,0) + F(0,1) x A(0,1) + ... + F(0,6) x A(0,6) + F(0,7) x A(0,1)
+ F(1,0) x A(1,0) + F(1,1) x A(1,1) + ... + F(1,6) x A(1,6) + F(1,7) x A(1,1)
+ .....
+ F(7,0) x A(7,0) + F(7,1) x A(7,1) + ... + F(7,6) x A(7,6) + F(7,7) x A(7,1)

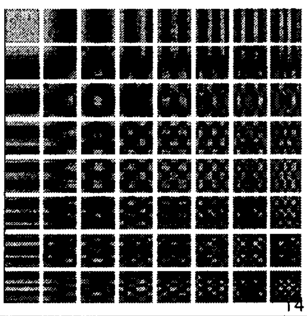
$$F(u,v) = \frac{2}{N}C(u)C(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)\cos\frac{(2x+1)u\pi}{2N}\cos\frac{(2y+1)v\pi}{2N}$$

$C(s) = 1/\sqrt{2}$ if s = 0, otherwise 1

$$f(x,y) = \frac{2}{N}\sum_{u=0}^{N-1}\sum_{v=0}^{N-1} C(u)C(v)F(u,v)\cos\frac{(2x+1)u\pi}{2N}\cos\frac{(2y+1)v\pi}{2N}$$

- **8x8 DCT changes 64 pixel values of f(x,y) to 64 DCT coefficients of F(u,v).**

- **What do we gain by "transform" ?**
  - **HVS says that, low freq. Component is more important than high freq. !!**
  - **Different quantization scale for different frequency components!**



Example of DCT

Relatively Smooth Area

186	187	185	183	185	187	186	186
186	185	187	187	186	186	186	188
191	187	190	187	186	187	190	185
188	187	190	186	187	188	184	187
185	186	185	189	189	189	189	184
185	185	184	188	185	187	185	190
182	187	185	185	187	185	183	191
185	188	187	186	187	186	183	188

1492	-1.26	-0.02	-1.38	0.50	-2.34	1.90	-1.09
1.43	2.21	1.56	2.90	-2.65	4.33	0.57	2.36
-4.02	-0.63	2.49	-2.10	0.41	-3.92	-0.12	-0.34
-3.47	-4.54	-0.26	0.89	1.02	-1.02	-1.83	-2.84
0.00	1.81	-1.90	2.50	-2.00	-1.32	-0.02	0.35
-0.90	0.98	2.40	-1.19	-0.61	-2.97	1.76	0.75
1.40	1.69	1.88	1.66	-0.98	2.85	0.01	-0.16
1.12	-0.30	-1.13	2.38	-1.98	1.75	-3.27	1.37

Highly Textured Area

138	168	103	207	142	154	184	160
119	185	191	195	164	111	153	99
221	129	165	98	132	223	239	169
220	88	166	228	68	174	146	185
137	129	183	172	178	143	161	158
165	113	203	194	131	215	164	159
158	157	128	165	212	63	165	160
142	154	150	191	167	191	151	170

Spatial Pixel Data

1281	-18.04	-18.93	-13.59	19.38	83.69	6.89	-17.23
-5.45	9.64	21.03	10.05	-39.04	29.92	-19.18	-21.01
-14.08	3.00	-34.18	-3.75	5.01	-62.78	-98.37	-11.11
-16.05	1.08	-50.22	-35.95	16.49	-14.95	-33.70	-3.21
-4.63	-19.75	-6.35	-16.82	35.13	12.26	14.40	-66.92
-4.29	-14.51	69.25	19.67	55.15	35.73	-6.86	-17.07
50.61	-59.71	41.63	67.51	-55.52	27.88	78.43	-34.99
-0.68	-30.89	27.75	35.73	-5.68	-49.74	-55.50	73.58

DCT Coefficients

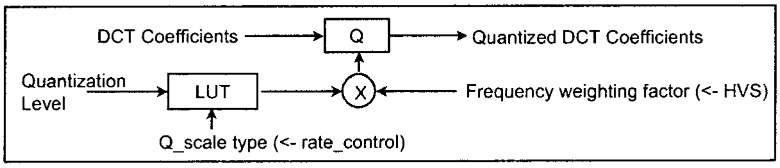
from Mobile & Calendar

Lecture Note by B. Jeon(Nov.23,2001)

15

Quantization

- **Mapping input data into finite number of values**
- **Why Quantization ? Why in transform domain ???**
- **Coarseness of quantizer**
  - **- quantization level**
  - **- frequency weighting factor (except intra DC)**



8	16	19	22	26	27	29	34
16	16	22	24	27	29	34	37
19	22	26	27	29	34	34	38
22	22	26	27	29	34	37	40
22	26	27	29	32	35	40	48
26	27	29	32	35	40	48	58
26	27	29	34	38	46	56	69
27	29	35	38	46	56	69	83

intra block

16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16

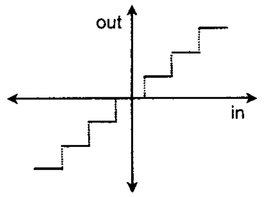
inter block

Lecture Note by B. Jeon(Nov.23,2001)

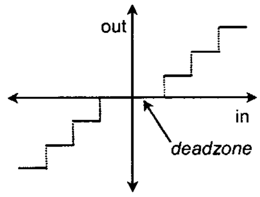
16

Quantizer Characteristics

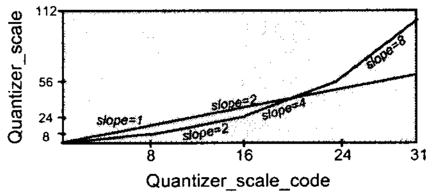
intra pictures



inter pictures



- Linear vs. piecewise linear quantizer (q_scale_type=0 / 1)



Example of DCT+Quantization

	-0.37	-0.04	0.07	0.41	0.38	-0.31	-0.11	-0.23	0	0	0	0	0	0	-1	0
	-0.45	-0.36	0.03	0.05	-0.04	-0.08	-0.18	-0.01	0	0	-1	0	0	0	1	0
	-0.08	0.00	-0.18	0.25	0.01	0.22	-0.37	-0.11	0	0	0	0	0	0	0	0
	-0.05	0.08	-0.22	0.05	0.49	0.05	0.30	-0.21	0	0	0	-1	0	0	0	0
Qstep = 1	0.37	0.25	-0.35	0.18	0.13	0.26	0.40	0.08	0	0	0	0	0	1	0	0
	-0.29	0.49	0.25	-0.33	0.15	-0.27	0.14	-0.07	0	0	0	0	0	0	-1	0
	-0.39	0.29	-0.37	-0.49	0.48	-0.12	0.43	0.01	0	0	0	0	0	0	0	0
	0.32	0.11	-0.25	-0.27	0.32	0.26	0.50	-0.42	0	0	0	0	0	0	0	0

	1.63	5.96	5.07	2.41	3.38	3.69	6.89	6.77	26	-9	8	-7	2	1	2	3
	2.55	1.64	5.03	2.05	0.96	5.92	4.82	2.99	-7	2	-1	-1	-4	-1	-3	1
	1.92	3.00	5.82	4.25	5.01	1.22	5.63	4.89	4	-2	2	1	-2	-2	2	-1
Qstep = 4	7.95	1.08	5.78	4.05	0.49	1.05	6.30	4.79	-1	-2	0	-1	1	-3	-2	-2
	3.37	4.25	1.65	7.18	3.13	4.26	6.40	5.08	4	0	3	1	0	3	2	-1
	3.71	1.49	5.25	3.67	7.15	3.73	1.14	6.93	0	1	4	-3	0	1	-2	0
	2.61	4.29	1.63	3.51	0.48	3.88	6.43	5.01	3	0	-2	0	3	-4	1	4
	7.32	1.11	3.75	3.73	2.32	6.26	0.50	1.58	2	-2	1	-2	1	-1	-5	-3

DCT Coefficient Error ***Spatial Pixel Error***

Question

- *Now, we have DCT+Quantized each block (or macroblock).*

DC	10	0	-2	0	0	0	0
0	-8	0	2	0	0	0	0
0	0	0	0	0	0	0	0
0	-2	0	0	0	0	0	0
0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

- ***We need to send this information.***
- ***How ?***

Variable-Length Coding : Scanning

- ***Quantized DCT coefficients are scanned to form (Run, Level) symbols.***
- ***Huffman codewords are assigned to represent (run,level) symbols.***

DC	1	5	6	14	15	27	28
2	4	7	13	16	26	29	42
3	8	12	17	25	30	41	43
9	11	18	24	31	40	44	53
10	19	23	32	39	45	52	54
20	22	33	38	46	51	55	60
21	34	37	47	50	56	59	61
35	36	48	49	57	58	62	63

Zig-zag Scan Order

DC	4	6	20	22	36	38	52
1	5	7	21	23	37	39	53
2	8	19	24	34	40	50	54
3	9	18	25	35	41	51	55
10	17	26	30	42	46	56	60
11	16	27	31	43	47	57	61
12	15	28	32	44	48	58	62
13	14	29	33	45	49	59	63

Alternate Scan Order

- (Example)
  - (0,10), (2, -8), (1,-2), (4,-2), (1,2), (5,1), EOB
- ***EOB (End of Block)*** is frequent symbol.
  - - intra picture
  - - inter picture
- Why alternate scan effective ?
  - "***alternate_scan***" in picture_coding_extention()

DC	10	0	-2	0	0	0	0
0	-8	0	2	0	0	0	0
0	0	0	0	0	0	0	0
0	-2	0	0	0	0	0	0
0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Variable-Length Coding : Concept of Huffman Coding

- (Michael Jackson's Bad)
  - Because I'm bad, I'm bad -- come on. Bad, bad -- really, really bad.
  - You know I'm bad, I'm bad -- You know it. Bad, bad -- really, really bad.
  - You know I'm bad, I'm bad -- come on, you know.
  - Bad, bad -- really, really bad

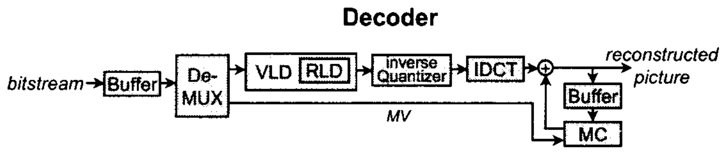
Symbols	Frequency	Codelength	Codeword
Because	1	5	00001
I'm	6	3	011
Bad	15	1	1
Come on	2	4	0001
It	1	5	00000
Really	6	3	010
You know	4	3	001

- ***Without entropy coding*** :
  - 35 (total symbols) x 3 bits = 105 bits
- ***With entropy coding*** :
  - 1x5 + 6x3 + 15x1 + 2x4 + 1x5 + 6x3 + 4x3 = 81 bits

Example of VLC Table (MPEG-2 Inter VLC)

Run	Level	Codeword
EOB		10
0	1	1s (first)
0	1	11s
0	2	0100s
0	3	0010 1s
...	...	...
1	1	011s
1	2	0001 10s
1	3	0010 0101s
1	4	0000 0011 00s
....	...	...
11	1	0010 0011s
12	1	0010 0010s
...	...	...
25	1	0000 0000 1110 0s
26	1	0000 0000 1101 1s

수신측(Decoder)의 역할

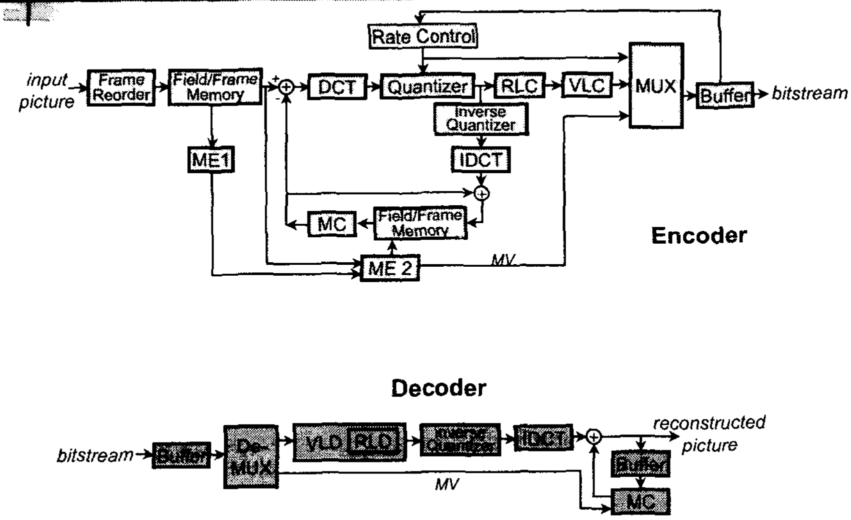


Lecture Note by B. Jeon(Nov.23,2001)

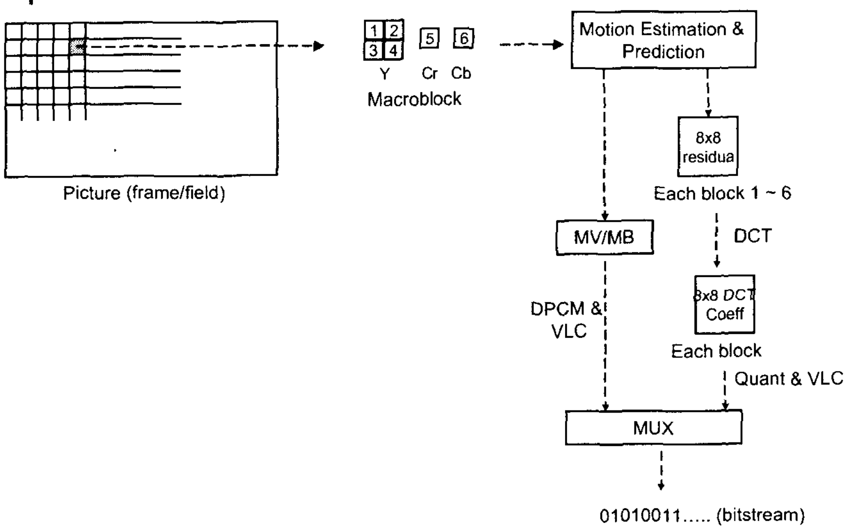
23

Concept of MPEG Inter Coding

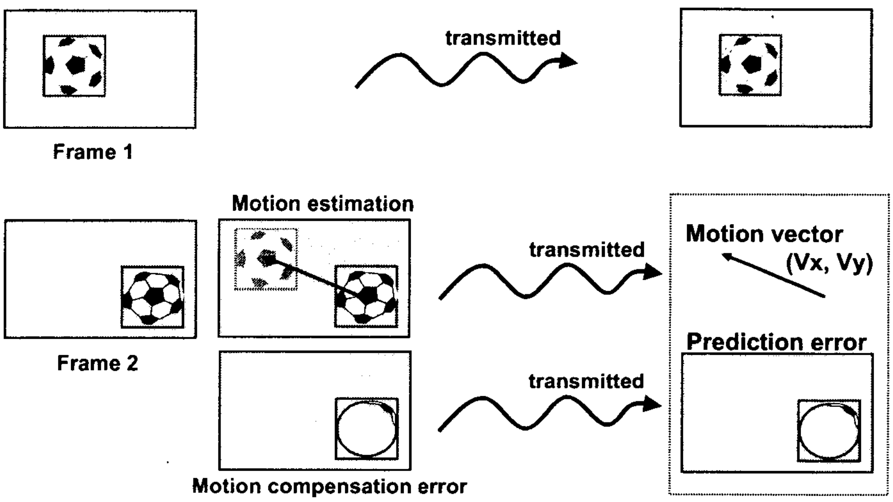
Inter Coding of MPEG-2



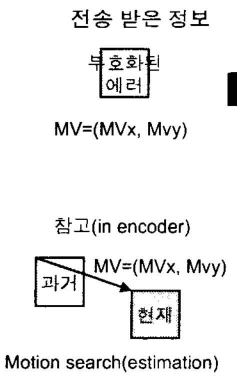
Procedure of Inter Coding

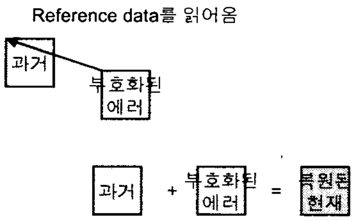


Concept of Motion Compensation

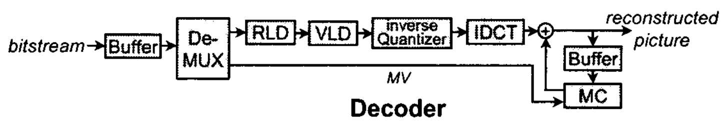


Motion-Compensated Prediction: Decoding ???



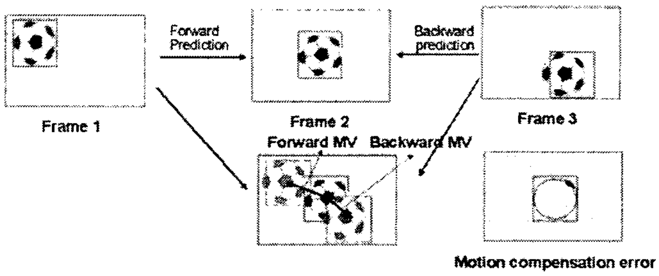


- ***The decoder must store the previous picture** while reconstructing the current one.*



Bi-directional Compensation

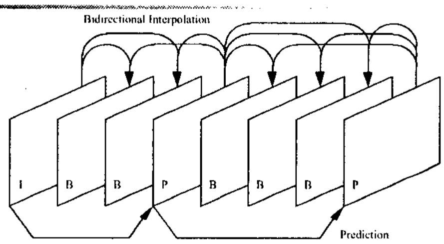
Forward/ Backward/ Bi-directional Prediction



- ***Picture Coding Type***:
  - I(intra), P(Predictive), B(Bi-directional interpolative) pictures
- M and N : N (number of frames in a GOP)
  - M (distance between successive reference pictures

(example) N = 15, M = 3

Display & Transmission Order



- Frame buffers for reference frames required in decoder
- ***Frame re-ordering necessary if B-frames exist.***

(ex) At the encoder input,

1	2	3	4	5	6	7	8	9	10	11	12	13
I	B	B	P	B	B	P	B	B	I	B	B	P

At the encoder output, in the coded bitstream, and at the decoder input,

1	4	2	3	7	5	6	10	8	9	13	11	12
I	P	B	B	P	B	B	I	B	B	P	B	B

At the decoder output,

1 2 3 4 5 6 7 8 9 10 11 12 13

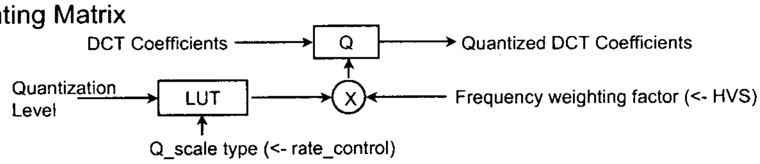
Comparison of Picture Types

	I-Picture	P-Picture	B-Picture
Coding Method	Intraframe : DCT-based	Hybrid Interframe: MCP Intraframe : DCT-based	Hybrid Interframe: MCI Intraframe : DCT-based
Compression	Less than 4	Less than 8	More than 8
Error Propagation	reset	error propagation	no propagation
functionalities	fast playback random access low delay coding	fast playback low delay coding	High compression

MCP : Motion Compensated Prediction *MCI : Motion Compensated Interpolation*

Quantization in motion compensated picture

- Weighting Matrix



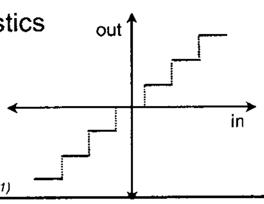
8	16	19	22	26	27	29	34
16	16	22	24	27	29	34	37
19	22	26	27	29	34	34	38
22	22	26	27	29	34	37	40
22	26	27	29	32	35	40	48
26	27	29	32	35	40	48	58
26	27	29	34	38	46	56	69
27	29	35	38	46	56	69	83

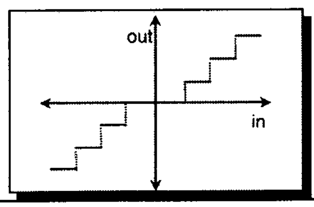
intra block

16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16

inter block

- Quantizer Characteristics





Macroblock Type in I, P, & B Picture

- ***Macroblock type in I-Picture***
  - intra
  - intra, Quant
- ***Macroblock Type in P Picture***
  - *MC (non zero MV)*
    - *coded*
      - ***coded***
      - ***coded, quant***
    - ***not coded***
  - *no MC (zero MV)*
    - ***coded***
    - ***coded, quant***
  - *intra*
    - ***coded***
    - ***coded, quant***
- ***Macroblock Type in B Picture***
  - *forward MC* / *backward MC* / *interpolative MC*
    - *coded*
      - ***coded***
      - ***coded, quant***
    - ***not coded***
  - *intra*
    - ***coded***
    - ***coded, quant***

Profile & Level

LEVELS		PROFILES				
	Layer	Simple	Main	SNR	Spatial	High
High	Enhancement		1920x1152x60			1920x1152x60
	Lower		NA			960x576x30
H1440	Enhancement		1440x1152x60		1440x1152x60	1440x1152x60
	Lower		NA		720x576x30	720x576x30
Main	Enhancement	720x576x30	720x576x30	720x576x30		720x576x30
	Lower	NA	NA	NA		352x288x30
Low	Enhancement		352x288x30	352x288x30		
	Lower		NA	NA		

MPEG-2 Summary

- Basic coding mode : interframe DCT with I, P, B pictures
- New field/frame prediction for interlaced
- For improved quality over MPEG-1,
  - improved quantization
  - new intra VLC
  - field/frame DCT
  - piecewise linear quantizer
- Better display over MPEG-1
  - syntax to facilitate 3:2 pulldown
  - pan and scan vector upto 1/16 resolution
- Scalability extension for hierarchical service, robustness, etc
  - spatial scalability
  - temporal scalability
  - SNR scalability
  - data partitioning

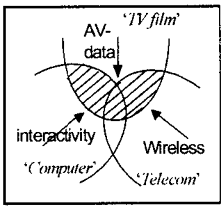
MPEG-4 Video

Why MPEG-4?
What is different?
Content-based?
Shape Coding
Motion Coding
Texture Coding

A bit of history on MPEG-4

- Participants of MPEG: Consumer Electronics, Computer Industry, Telecom Operators, Academia.
- 1993 : started as 'Very Low Bitrate Audio-visual Coding'.
- 1994: goal changed to '**Coding of Audio-visual Objects** '.
- And finally MPEG-4 attempts to provide a bridge between the www and conventional AV media
- **1999/2000: MPEG-4 Standard**
  - *Flexible Multimedia Communications* (5 kbits/s - 50 Mbit/s)

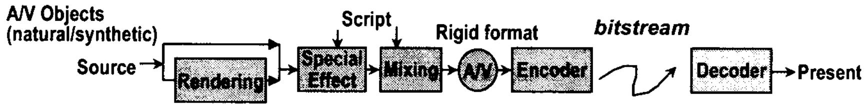
From the MPEG-4 Call for Proposals (drafted 1994)



What's different from the old..(1)

- *Content-based Interactivity*
  - **Manipulation of Scene Content on Bitstream Level (~ scalability)**
  - **Content-based Manipulation and Bitstream Editing**
  - **Hybrid Natural and Synthetic Data Coding**
  - **Improved Temporal Random Access**
- *Compression*
  - **Improved Coding Efficiency**
  - **Bitrate: 5 kbps ~ 5 Mbps; Resolution : small ~ TV**
  - **Support progressive & interlaced**
  - **Coding of Multiple Concurrent Data Streams**
- *Universal access*
  - **Error resilience/robustness in Error-prone (ex; mobile) Environments**
  - **Content-based Scalability**

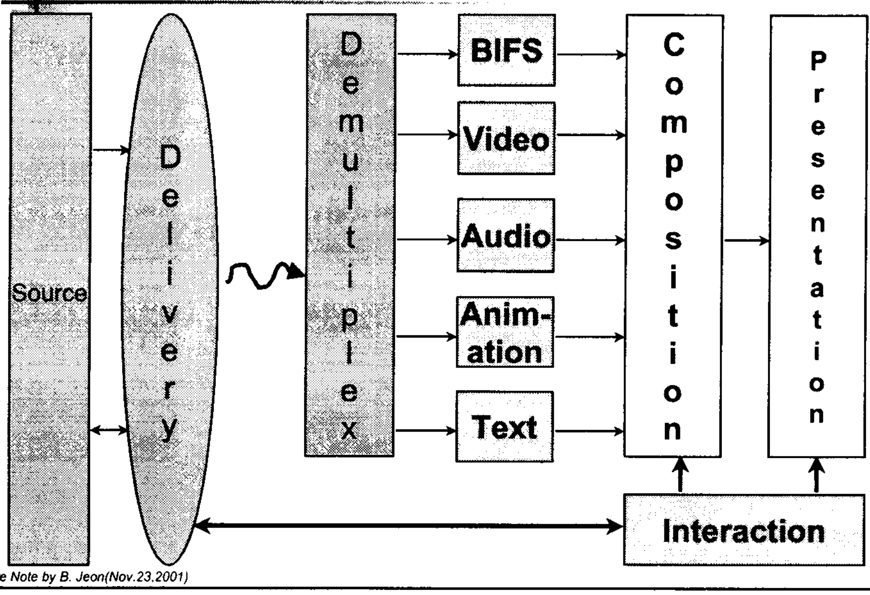
What's different from the old..(2)



Conventional Approach

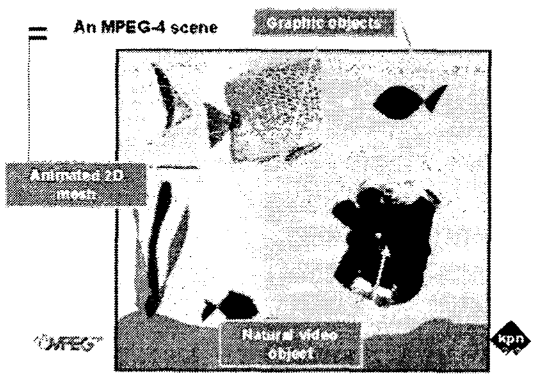
	MPEG-1	MPEG-2
System :	ISO/IEC 11172-1	ISO/IEC 13818-1
Video :	ISO/IEC 11172-2	ISO/IEC 13818-2
Audio :	ISO/IEC 11172-3	ISO/IEC 13818-3
Interaction:		ISO/IEC 13818-6

MPEG-4 = More Interactivity to user !!!



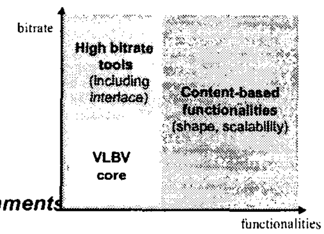
MPEG-4 Scene Example

- An example scene supported by MPEG-4 (*not simple profile*)



from the MPEG official website

MPEG-4 Functionalities



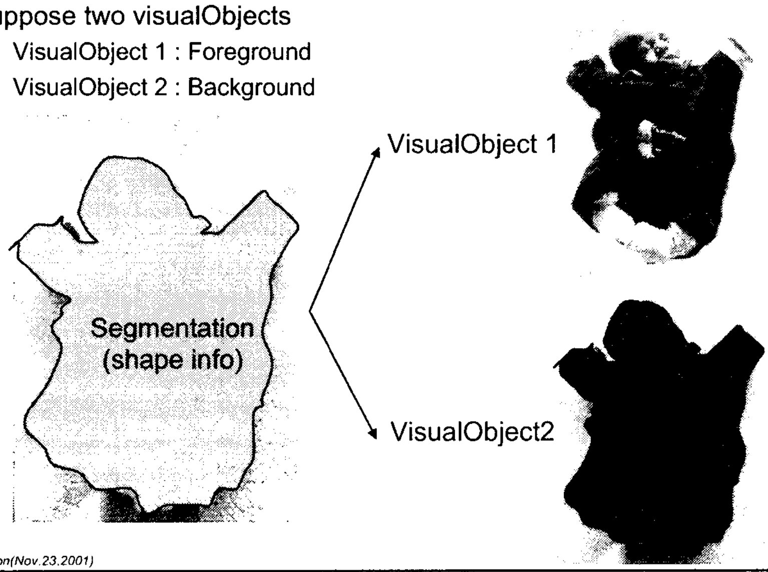
- ***Baseline***
  - ***Coding Efficiency***
    - ***5 kbits/s - 50 Mb/s***
    - ***Resolution: Small – TV***
    - ***Progressive/Interlace***
  - ***Error Resilience/Robustness: Mobile Environments***
  - ***Scalability (Spatial/Temporal)***
- ***Extended***
  - ***Arbitrarily-Shaped Objects in a Coded Scene***
  - ***High Interaction With Scene Content***
  - ***Manipulation of Scene Content on Bitstream Level***

Major Natural Video Tools

- Binary shape
- Padding
- Motion compensation
  - Overlapped motion compensation
  - Advanced motion compensation
- Quantization
  - Method 1
  - Method 2
  - Non-linear
- AC/DC prediction
- Scanning
- I, P, B modes
- Temporal scalability
  - Type 1
  - Type 2
- Spatial Scalability
- Error resilience
  - Slice synchronization
  - Extended header code
  - Data partitioning
  - Reversible VLC
- Static sprites
  - Basic
  - Low delay
  - Scalable
- Interlaced coding
- 12-bit video
- Static texture

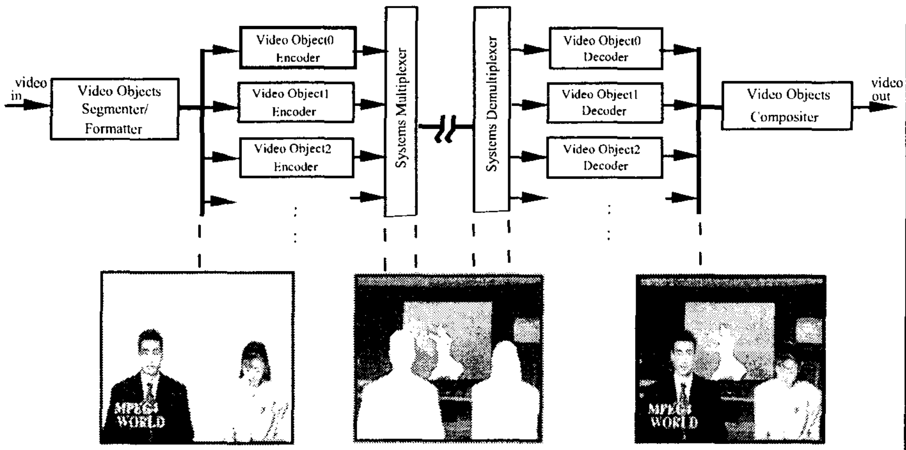
Object Extraction

- Suppose two visualObjects
  - VisualObject 1 : Foreground
  - VisualObject 2 : Background



Content-Based Coding of Video

- ***Each Video Object*** in a Scene is Coded and Transmitted **Separately**

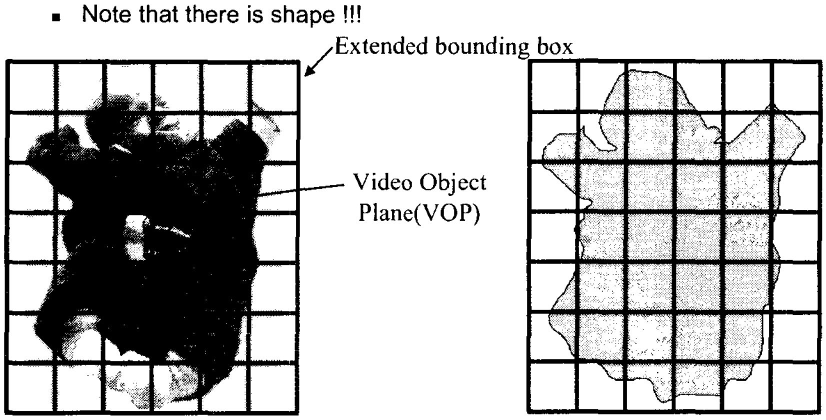


Lecture Note by B. Jeon(Nov.23,2001)

8 45

Coding of VOP

- Once VOP is found, encode VOP on macroblock basis!!!
  - Note that there is shape !!!

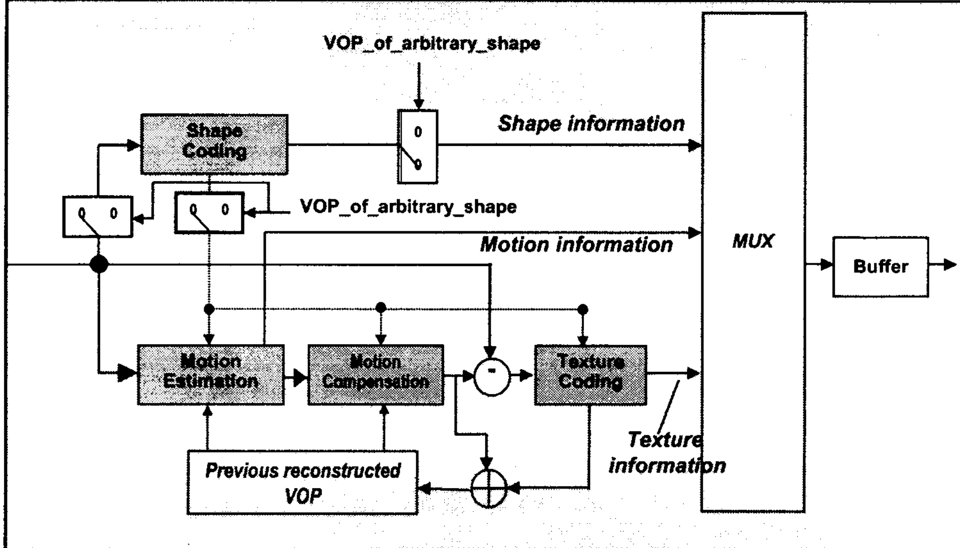


Texture Information ***Motion Information*** ***Shape Information***

Lecture Note by B. Jeon(Nov.23,2001)

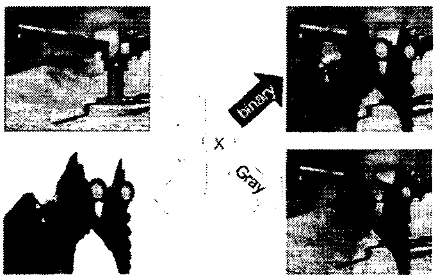
46

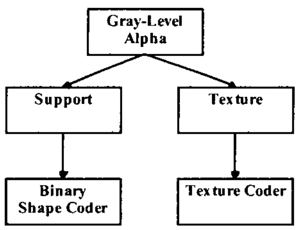
VOP Encoder



Coding of Shape Information

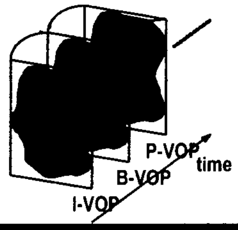
- Alpha Plane: Binary vs. Gray alpha plane
  - Alpha plane is partitioned into 16X16 blocks(alpha block)
  - Encoding/decoding is done per ***alpha block***
- Binary alpha plane : encoded by ***modified CAE***
- Greyscale alpha plane : encoded by ***motion compensated DCT***
  - Support fuction(*shape* information) ; binary shape coding
  - Alpha values(*transparency* information) ; texture coding

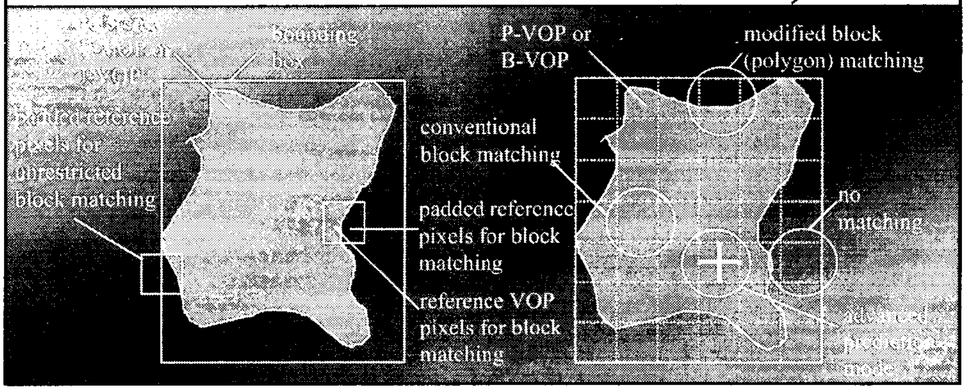




Motion Estimation & Compensation

- Motion compensated coding modes(I, B, P)





Modified Block(Polygon) Matching

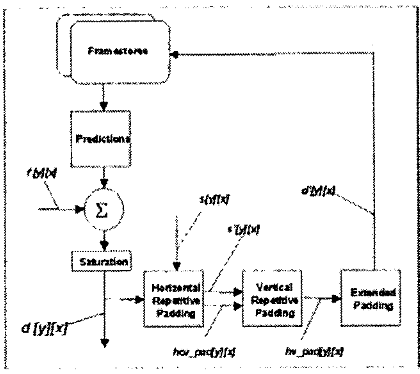
- Polygon matching for an arbitrary shape of VOP:
  - SAD evaluated excluding exterior of current MB.

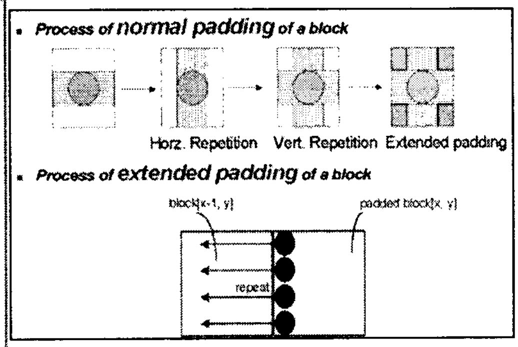
$$SAD_N(x, y) = \sum_{i=1, j=1}^{N,N} |original - previous| * (!(Alpha_{original} == 0)),$$

$x, y =$ "up to [-64, 63]", N = 16 or 8

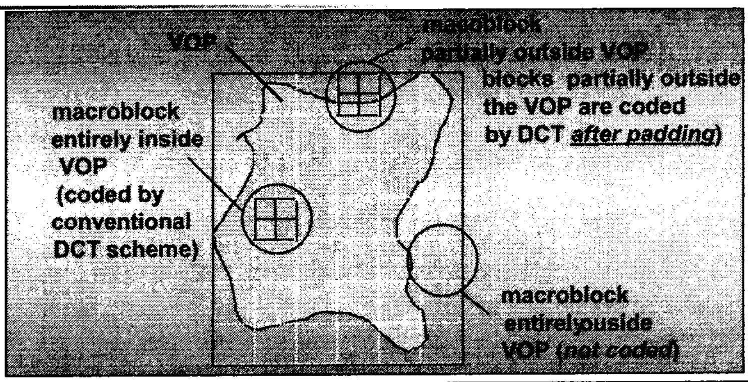
Padding Process

- Padding defines the value of luminance and chrominance samples outside of the VOP for prediction of arbitrarily shaped object.





Texture Coding in Intra/Inter MBs



Intra block(8*8)			Residual block of inter block (8*8)
Interior block	Boundary block	Exterior block	
Identical to H.263 texture coding	LPE padding followed by H.263 texture coding	Not coded	Zero padding followed by H.263 texture coding

Low Pass Extrapolation (LPE) Padding

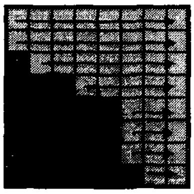
- Purpose : To enhance energy packing efficiency of DCT in ***intra*** MB.
- Low Pass Extrapolation Padding (LPE) Procedure
  - **Step 1 : Calculate mean value m**

$$m = \frac{1}{N} \sum_{(i,j)\in R} f(i,j) \qquad f(i,j): \text{pixel value at (i,j) inside object R}$$

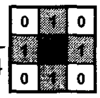
  - **Step 2 : Assign *m* to each block pixel outside of R** $f(i,j) = m, \ \textit{for all pixel} \notin R$
  - **Step3 : Low pass filtering to each outer pixel of R from top-left to btm-right.**

$$f(i,j) = [f(i,j-1) + f(i-1,j) + f(i,j+1) + f(i+1,j)]/4$$

- division is done by rounding to nearest integer.
- If referring to outside pixel, use 0 and deduce the number 4 accordingly.



■ Object pixel
■ Padded pixel
→ Order of LPE

$\frac{1}{4}$ 

Discrete cosine transform

- N x N two dimensional DCT

$$F(u,v) = \frac{2}{N} C(u)C(v) \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y) \cos\frac{(2x+1)u\pi}{2N} \cos\frac{(2y+1)v\pi}{2N}$$

With u, v, x, y = 0, 1, 2, ... N-1
x, y : spatial coordinates in the sample domain
u, v : coordinates in the transform domain

$$C(u), C(v) = \begin{cases} \frac{1}{\sqrt{2}} & \textit{for } u,v=0 \\ 1 & \textit{otherwise} \end{cases}$$

- Inverse DCT(IDCT)

$$f(x,y) = \frac{2}{N} \sum_{u=0}^{N-1}\sum_{v=0}^{N-1} C(u)C(v)F(u,v) \cos\frac{(2x+1)u\pi}{2N} \cos\frac{(2y+1)v\pi}{2N}$$

- f(x,y) is a real number.
  - integer-number IDCT : $f'(x,y) = round(f(x,y))$
  - saturate integer-number : $f''(x,y) = saturate(f'(x,y))$
  - saturate

$$saturate(x) = \begin{cases} -2^n & x < -2^n \\ x & -2^n \le x \le 2^n - 1 \\ 2^n - 1 & x > 2^n - 1 \end{cases}$$

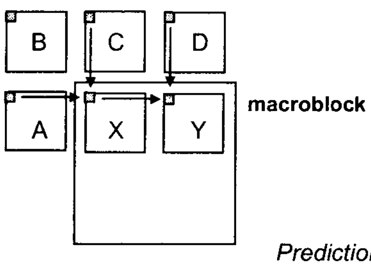
Where n is the number of bits per pixel.

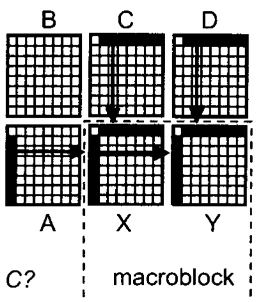
Quantization in MPEG-4

- Two quantization Methods are defined
  - Quant_type=1: (MPEG quantizer)
  - Quant_type=0: (H.263 quantizer)
- 'quant-type' field in videoObjectLayer()
- In simple profile: only 2nd quantizer(H.263 quantizer) is defined.

ACDC Prediction for TCOEFF

- ***Adaptive AC and DC prediction***



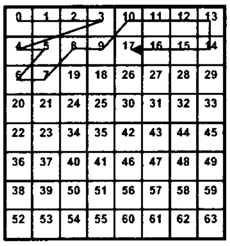


Prediction from A or C?

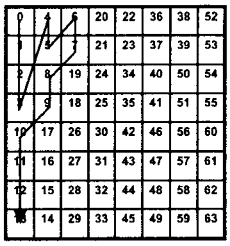
- ***Some Questions:***
  - **Why?**
  - **Why only for Intra block?**
  - **Why only for top row and first column?**
  - **DC predictive encoding by using DC gradient of adjacent block**

VLC encoding of Intra MB

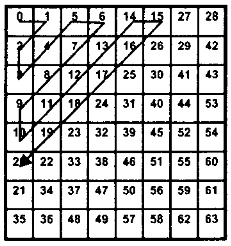
- Intra DC: VLC switching
- Intra AC: one scan among following threes
  - ACpred_flag = 0 (no ACDC prediction) : Zigzag scan always
  - ACpred_flag = 1:
    - **Horizon adjacent block prediction : Vertical scan**
    - **Vertical adjacent block prediction : Horizontal scan**



Horizontal scan



Vertical scan

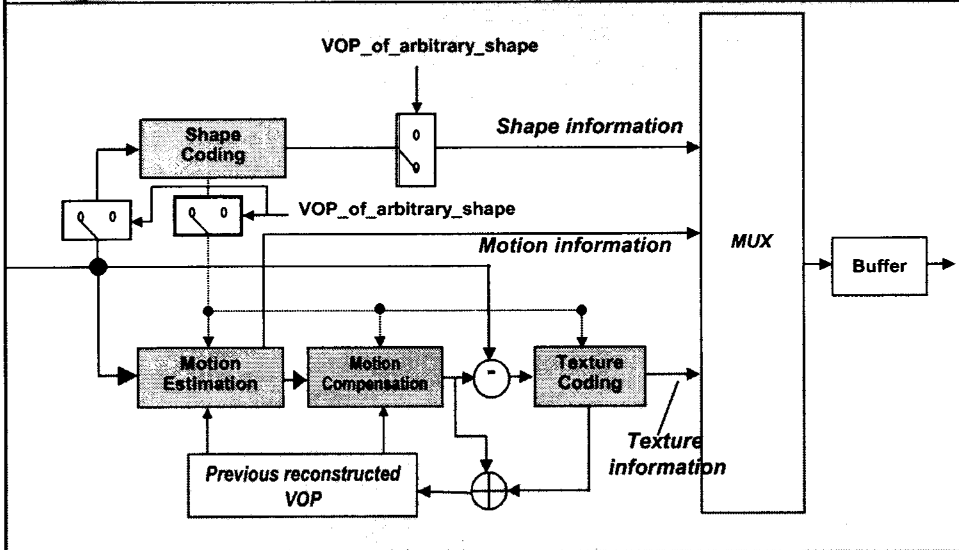


Zigzag scan

VLC Encoding of TCOEFF

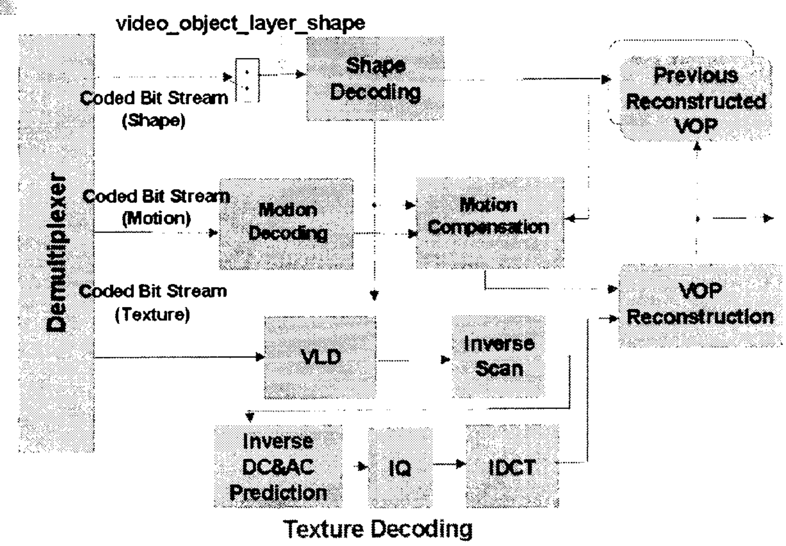
- ***Scan direction as specified by ACDC prediction (Intra AC; Intra DC+AC)***
- ***Always zigzag scan for inter DC+AC***
- ***3D VLC code for symbol (LAST, RUN, LEVEL):***
  - LAST
    - 0 : There are more nonzero coefficients in the block.
    - 1 : This is the last nonzero coefficient in the block.
  - RUN : # zero coefficients proceeding the current nonzero coefficient.
  - LEVEL : Magnitude of non-zero coefficient.
- ***Escape coding:***
  - ESCAPE : 7 bit (0000 011)
  - LAST : 1 bit (0: Not last coefficient, 1: Last nonzero coefficient)
  - RUN : 6 bit
  - LEVEL : 8 bit
- Same VLC table for intra chrominance AC and intra AC luminance coefficients.

VOP Encoder



Lecture Note by B. Jeon(Nov.23,2001) 59

Block Diagram of Natural Video Decoding



Lecture Note by B. Jeon(Nov.23,2001) 60