EE 6886: Topics in Signal Processing -- Multimedia Security System

Lecture 2: Multimedia Compression Techniques and Standards

Ching-Yung Lin
IBM T. J. Watson Research Center
Hawthorne, NY 10532, USA

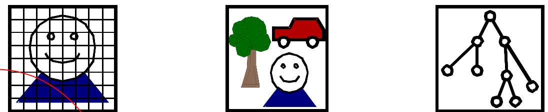
Course Outline

- Multimedia Security :
  - Multimedia Standards – Ubiquitous MM
  - Encryption and Key Management – Confidential MM
  - Watermarking – Uninfringible MM
  - Authentication – Trustworthy MM
- Security Applications of Multimedia:
  - Audio-Visual Person Identification – Access Control, Identifying Suspects
  - Surveillance Applications – Abnormality Detection
  - Media Sensor Networks – Event Understanding, Information Aggregation

Multimedia Standards: Towards Knowledge Management and Transaction Enrichment for Digital Media

Signals → Features → Semantics → Knowledge

Recent past ——————————————→ Near future



Focus of this lecture

Applications			
MPEG-1,-2,-4 Video storage Broadband Streaming video delivery	**MPEG-4,-7** Content-based retrieval Multimedia filtering Content adaptation	**MPEG-7** Semantic-based retrieval and filtering Enterprise content mgmt.	**MPEG-21** E-commerce of Electronic content Digital items
Problems and Innovations			
Compression Coding Communications	Similarity searching Object- and feature-based coding	Modeling and classification Personalization and summarization	Media mining Decision support IPMP (rights)

MPEG - 1, - 2 | MPEG - 4 | MPEG - 7 | MPEG - 21

Information in Multimedia

- Fact 1: A film frame = 480 x 260 pixels = 374,400 bytes = 2,995,200 bits = 599,040 alphabets = 124,025 English words
- Fact 2: A 2 hour 10 mins movie = 187,200 frames = 70,087,680,000 bytes = 23,217,480,000 English words = 7,067 years of words a person can talk
- Fact 3: 128 hours of video = 13,648,457 frames = 31,131,584,804,571 bytes = storage space of 1556 20-GB hard disks at my laptop

Why needs compression?

- Demand:
  - VHS: 352 x 240 x 30 x 24 = 60.8 Mb / s
  - DVD: 720 x 480 x 30 x 24 = 248.8 Mb / s
  - HDTV: 1280 x 720 x 60 x 24 = 1327.2 Mb / s
- Supply:
  - Telephone: < 56 Kb/s
  - ISDN: 64 – 144 Kb/s
  - T1: 1.5 Mb/s
  - Ethernet: 10 - 100 Mb/s
  - 802.11b: 1 - 10 Mb/s
  - 802.11g: 50 - 100 Mb/s
  - CD-ROM: 1.2 Mb/s

Compression Types

- Lossless Compression
- Lossy Compression

Multimedia Compression

- Text Compression
  - Ancient Chinese

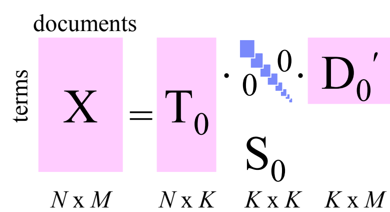
"When things are investigated, then true knowledge is achieved; when true knowledge is achieved, then the will becomes sincere; when the will becomes sincere, then the heart sees correctly; when the heart sees correctly, then the personal life is cultivated; when the personal life is cultivated, then the family life is regulated; when the family life is regulated, then the national life is orderly; and when the national life is orderly, then there is peace in this world."

From *Li-Ji* (Record of Rites)

誠意　正心　修身　齊家　治國　平天下

Multimedia Compression

- Document Indexing/Compression:
  - Latent Semantic Analysis(LSA) [Landauer, Dumais 1997]
    - Capture the concepts instead of words
      - Synonym
      - Polysemy
    - Truncated SVD: optimal least-square projection to reduce dimensionality



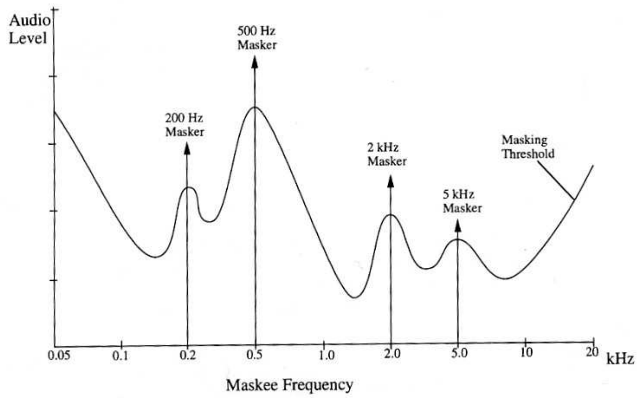
Example of text data: Titles of Some Technical Memos

c1: *Human* machine *interface* for ABC *computer* applications
c2: A *survey* of *user* opinion of *computer system response time*
c3: The *EPS user interface* management *system*
c4: *System* and *human system* engineering testing of *EPS*
c5: Relation of *user* perceived *response time* to error measurement

m1: The generation of random, binary, ordered *trees*
m2: The intersection *graph* of paths in *trees*
m3: *Graph minors* IV: Widths of *trees* and well-quasi-ordering
m4: *Graph minors*: A *survey*

Multimedia Compression

- Audio Compression



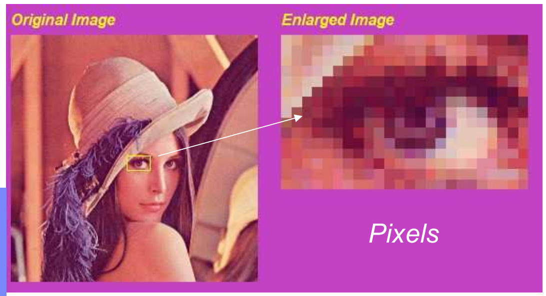
Inaudible Audio Masks

Multimedia Compression

- Image Compression
  - Graphics
  - Natural Images
- Video Compression
  - Animated Graphics
  - Motion Pictures
  - Streaming and Mobile Applications

Digital Image Representations

- Pixels
- Color Formats:
  - RGB
  - YUV
  - YCbCr



$Y = 0.299R + 0.587G + 0.114B$
$U = B – Y$
$V = R – Y$
$Cb = (U/2) + 0.5$
$Cr = (V/1.6) + 0.5$

$R,G,B = 0 \sim 1$

- Demo Examples (in Prof. Shih-Fu Chang's DIP course): http://www.ctr.columbia.edu/~sfchang/course/dip/demos/Classes.html

E 6886 Topics in Signal Processing: Multimedia Security Systems

Common Digital Image Formats

- JPEG
- GIF
- BMP
- TIFF
- PPM, PGM, PBM
- JPEG-2000

© 2006 Ching-Yung Lin, Dept. of Electrical Engineering, Columbia Univ.

BMP Format

- Internal image format in Windows system.
- Number of colors: 2, 16, 256, 16777216 (or extended versions: 16- bits and 32-bit colors)
- Raw (uncompressed) or lossless compressed (Run Length Encoding)
- Format:
  - Bitmap-file header: type, size, layout of a device-independent bitmap file
  - Bitmap-information header: dimensions, compression type, color format
  - Color Table: mapping table from RGB to color indices
  - Image data: row by row
    - Each pixel or color table is represented by B, G, R.
    - Each row has to be the multiple of 4 bytes.
- Reference: http://netghost.narod.ru/gff/graphics/summary/micbmp.htm

GIF Format

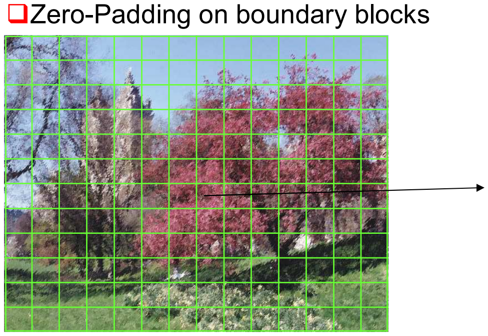
- Use LZW for lossless compression
- Limits to 256 colors
- Suitable for logo, graphics

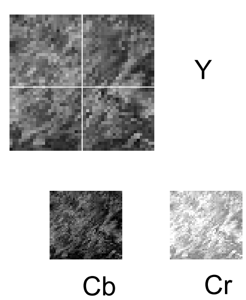
JPEG Format

- Divide Images into 8x8 (luminance) or 16x16 (chrominance) blocks
- DCT
- Quantization
- Run-level grouping
- VLC encoding

Macroblocks

- An image is divided into 8x8 blocks for the luminance components
- An image is divided into 16x16 blocks for the chrominance components
    - These chrominance blocks are down-sampled to 8x8 blocks (called 4:2:2 format)
- Zero-Padding on boundary blocks





Discrete Frequency-Domain Analysis

- Discrete Sine Transform
- Discrete Cosine Transform:
  - (Forward 8-coefficient) DCT

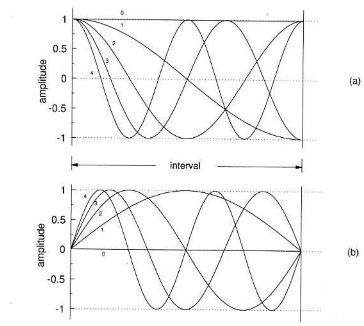
$$F(\mu) = \frac{C(\mu)}{2} \sum_{x=0}^{7} f(x) \cos[(2x+1)\mu\pi/16]$$

$$C(\mu) = 1/\sqrt{2}, if\ \mu = 0$$
$$C(\mu) = 1, if\ \mu > 0$$

  - Inverse 8-coefficient DCT

$$f(x) = \sum_{\mu=0}^{7} \frac{C(\mu)}{2} F(\mu) \cos[(2x+1)\mu\pi/16]$$

- Discrete Fourier Transform



(a) Continuous cosine bases;
(b) Continuous Sine bases

Discrete Frequency-Domain Analysis

- Discrete Sine Transform
- Discrete Cosine Transform:
  - (Forward 8-coefficient) DCT

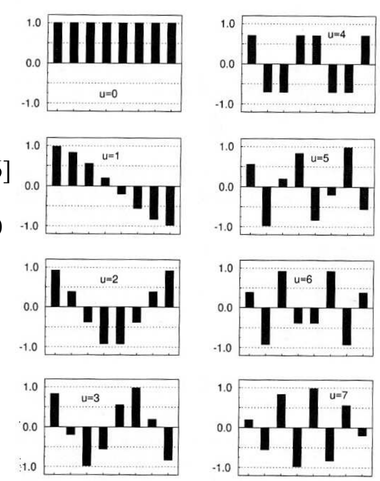
$$F(\mu) = \frac{C(\mu)}{2} \sum_{\mu=0}^{7} f(x) \cos[(2x+1)\mu\pi/16]$$

$$C(\mu) = 1/\sqrt{2}, if\ \mu = 0$$
$$C(\mu) = 1, if\ \mu > 0$$

  - Inverse 8-coefficient DCT

$$f(x) = \sum_{\mu=0}^{7} \frac{C(\mu)}{2} F(\mu) \cos[(2x+1)\mu\pi/16]$$

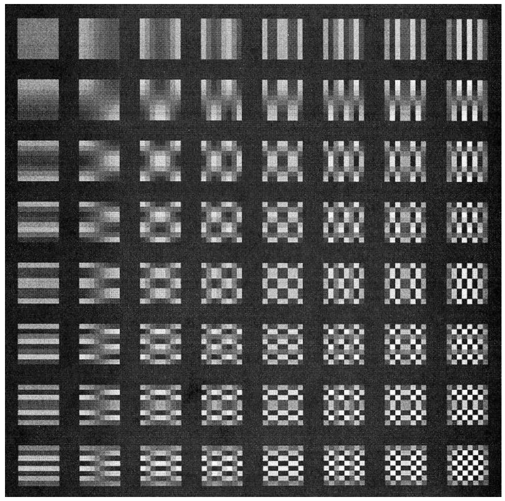
- Discrete Fourier Transform



DCT bases

Two-Dimensional Spatial Frequencies

- Spatial Frequencies decomposition



2D-DCT bases

Discrete Cosine Transform -- I

- Two-dimensional DCT:

$$F(\mu,\nu)=\frac{C(\mu)}{2}\frac{C(\nu)}{2}\sum_{y=0}^{7}\sum_{x=0}^{7}f(x,y)\cos[(2x+1)\mu\pi/16]\cos[(2y+1)\nu\pi/16]$$

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

$$C(\mu)=1, if\ \mu>0$$

Enlarged 8x8 Image

Original x[m][n]

158	165	158	158	158	158	155	158
155	161	155	150	155	155	158	151
150	155	155	150	161	161	174	174
171	167	171	159	151	151	134	117
96	94	90	102	102	108	108	101
108	108	96	108	108	108	108	108
108	108	103	108	108	120	110	117
110	117	117	123	125	129	124	127

DCT x[u][v]

1026	-4	-1	10	-9	0	-4	-6
160	14	6	-6	-4	0	-4	8
36	-18	5	-8	0	-9	0	-3
-86	-3	5	4	6	-2	-1	-6
-2	20	-13	2	-2	0	2	0
50	21	-5	-2	-2	-1	6	-3
15	-17	9	-6	7	0	-7	2
-36	-40	12	-4	6	-1	0	-1

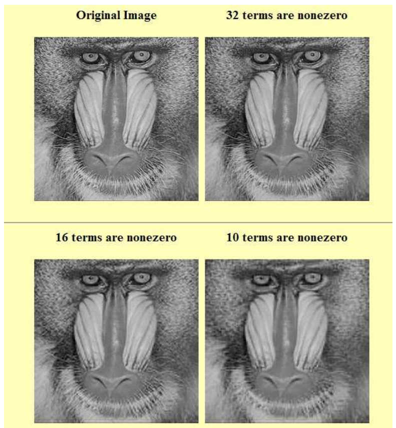
Discrete Cosine Transform -- II

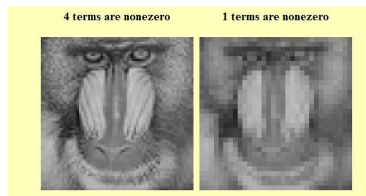
- Two-dimensional Inverse DCT:

$$f(x,y)=\sum_{\mu=0}^{7}\sum_{\nu=0}^{7}\frac{C(\mu)}{2}\frac{C(\nu)}{2}F(\mu,\nu)\cos[(2x+1)\mu\pi/16]\cos[(2y+1)\nu\pi/16]$$

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

$$C(\mu)=1, if\ \mu>0$$





Quantization

- Quantization Table

(a)

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

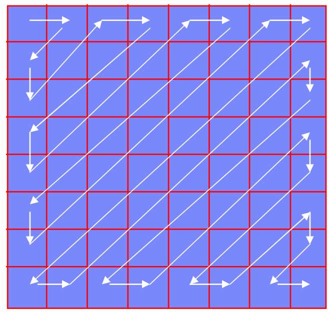
(b)

17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

Table : The quantization tables, $\mathbf{Q_{50}}$, of JPEG compression with *Quality Factor(QF) = 50* : (a) luminance,(b) chromnance. The quantization tables, $\mathbf{Q}_{QF}$ of other Quality Factor are $Integer\ Round(\mathbf{Q}_{50}\cdot q)$, where $q=2-0.02\cdot QF$, if $QF\geq 50$, and $q=\frac{50}{QF}$, if $QF<50$. In the baseline JPEG, $\mathbf{Q}_{QF}$ will be truncated to be within 1 to 255.

Zig-Zag Order

- An order to sort the 2D DCT coefficients to 1D signals



Run-Level Grouping (on AC coefficients)

- Convert 2D quantized DCT coefficients to (Run, Level) pairs
  - DPCM+VLC is used for DC coefficient coding: ΔDC = DC - P

96	12	-5	2	0	0	0	0
8	0	0	0	0	0	0	0
0	-1	2	0	0	0	0	0
1	0	1	0	0	0	0	0
5	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

(Run, Level):

(0,12), (0,8), (2,-5), (0,2), (1,-1), (0,1), (0,5), (1,2), (5,1), (4,-1), EOB

Variable Length Coding

- Usually, JPEG carries a VLC Table. MPEG uses a stand table.
- VLC: (Run, Level)
  (0,12):00000000110100
  (0,8):0000000111010
  (2,-5): 00000000101001
  (0,2): 01000
  (1,-1): 0111
  (0,1): 110
  (0,5): 001001100
  (1,2):0001100
  (5,1): 0001110
  (4,-1):001101
  EOB: 10

  ➔Total 84 bits

  If assume 6 bits for DC, then Compression ratio =
  (8x64)bits / 90 bits = 5.69

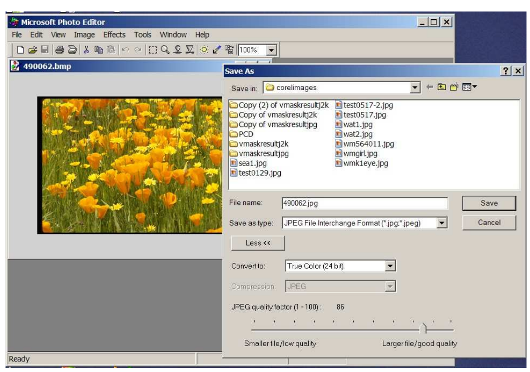
only for MPEG nonintra coding

run/level	VLC	bits
0/1	1s (first)	2
0/1	11s (next)	3
0/2	0100 s	5
0/3	0010 1s	6
0/4	0000 110s	8
0/5	0010 0110 s	9
0/6	0010 0001 s	9
0/7	0000 0010 10s	11
0/8	0000 0001 1101 s	13
0/9	0000 0001 1000 s	13
0/10	0000 0001 0011 s	13
0/11	0000 0001 0000 s	13
0/12	0000 0000 1101 0s	14
0/13	0000 0000 1100 1s	14
0/14	0000 0000 1100 0s	14
0/15	0000 0000 1011 1s	14
0/16	0000 0000 0111 11s	15
2/1	0101 s	5
2/2	0000 100s	8
2/3	0000 0010 11s	11
2/4	0000 0001 0100 s	13
2/5	0000 0000 1010 0s	14
3/1	0011 1s	6
3/2	0010 0100 s	9
3/3	0000 0001 1100 s	13
3/4	0000 0000 1001 1s	14
4/1	0011 0s	6
4/2	0000 0011 11s	11
4/3	0000 0001 0010 s	13
5/1	0001 11s	7
1/1	011s	4
1/2	0001 10s	7
1/3	0010 0101 s	9
1/4	0000 0011 00s	11
1/5	0000 0001 1011 s	13
1/6	0000 0000 1011 0s	14
1/7	0000 0000 1010 1s	14
1/8	0000 0000 0011 111s	16
1/9	0000 0000 0011 110s	16
26/1	0000 0000 1101 1s	14
27/1	0000 0000 0001 1111 s	17
28/1	0000 0000 0001 1110 s	17
29/1	0000 0000 0001 1101 s	17
30/1	0000 0000 0001 1100 s	17
31/1	0000 0000 0001 1011 s	17
End_of_block	10	2
Escape	0000 01	6

s: ‘0’ for positive, ‘1’ for negative

JPEG Examples

- Demo using MS Photo Editor



Discrete Fourier Transform

- (Forward) DFT:

$$F(\mu) = \frac{1}{N}\sum_{x=0}^{N-1} f(x)e^{-jx2\pi\mu/N}$$

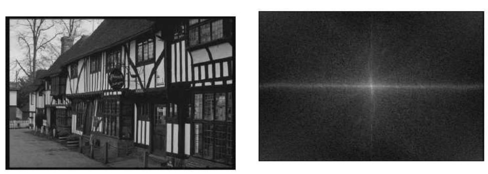
- Inverse DFT:

$$f(x) = \frac{1}{N}\sum_{\mu=0}^{N-1} F(\mu)e^{jx2\pi\mu/N}$$

➔What are the relationships between DST, DCT, and DFT?
➔What are the relationships between DFT and FT?

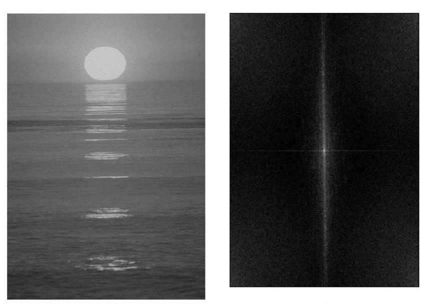
DFT Spectrums

- Image Examples



(a) (b)

Image with dominant vertical structure and its DFT



(a) (b)

Image with dominant horizontal structure and its DFT.

Continuous Fourier coefficients of continuous images after RSC

- Rotation in time => Rotation in frequency

$$x_R(t_1,t_2) = x(t_1\cos\theta - t_2\sin\theta, t_1\sin\theta + t_2\cos\theta) \overset{F}{\longleftrightarrow}$$
$$X(f_1\cos\theta - f_2\sin\theta, f_1\sin\theta + f_2\cos\theta) = X_R(f_1,f_2)$$

- Scaling in time => Scaling in frequency

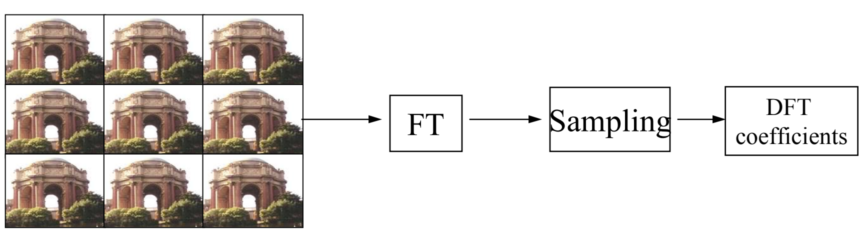
$$x_S(t_1,t_2) = x(\tfrac{t_1}{\lambda_1},\tfrac{t_2}{\lambda_2}) \overset{F}{\longleftrightarrow} X(\lambda_1 f_1, \lambda_2 f_2) = X_S(f_1,f_2)$$

- Translation in time => Phase shift in frequency
- Information Loss in time => Noise in frequency
- Change of Image Size in time => No definition in frequency

E 6886 Topics in Signal Processing: Multimedia Security Systems

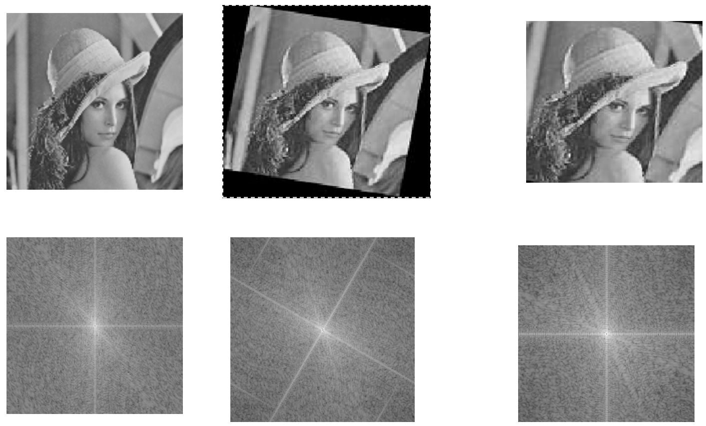
Relationships between DFT coefficients and Fourier coefficients

- DFT: Samples of the Fourier coefficients of the repeated discrete image.



- Discrete Spectrum Analysis:
  - Periodic signal in the time/spatial domain -> Discrete samples in the frequency domain
  - Discrete samples in the time/spatial domain -> Periodic signals in the frequency domain
- DFT is based on the assumption of repeating the signals over time/space, then take samples in the frequency domain of the Fourier Transform Values

DFT coefficients after rotation



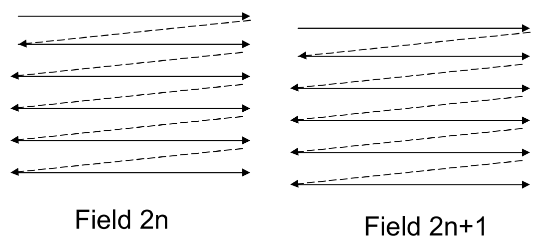
- Characteristics:
  - "cross" effect, Cartesian sampling points

JPEG - 2000

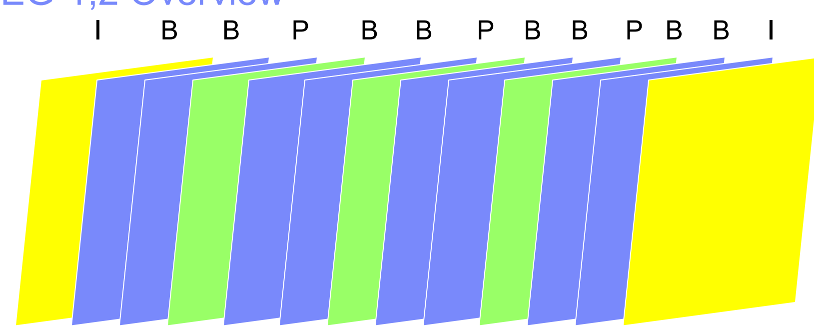
- Based on Wavelet bases
- Support Scalable Coding:
  - decoding can be terminated at any point.
  - Coarse-image is transmitted first, then more and more details coming.

Video Representation

- Raw Data Amount:
  - NTSC Analog Video: 525 lines, 480 visible, interlaced scan, 59.94Hz (~ 352x240 MPEG-1 format)
  - DVD: 720 x 480 x 30 x 24 = 31.1 MB / s
  - HDTV: 1280 x 720 x 60 x 24 = 165.9 MB / s
- Interlaced Video Scanning Patterns



MPEG-1,2 Overview



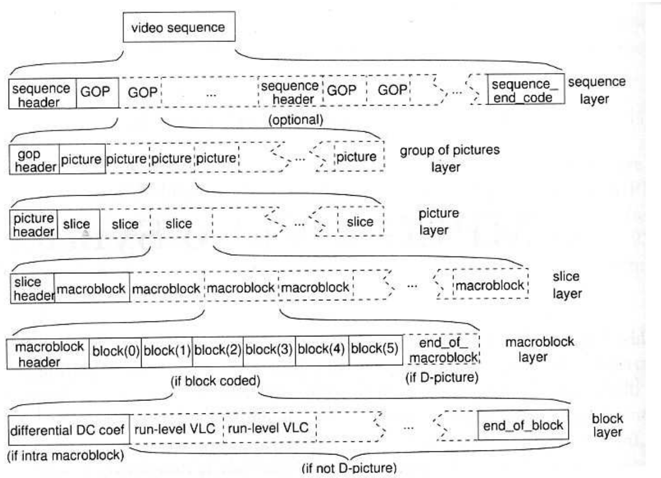
- Intraframe: I frames
- Interframe: P and B frames

- MPEG-1: 352x240 or 352x264 – for VCD
- MPEG-2: (1) multiple resolutions, e.g., 1024x768 – for compatibility with TV. (2) field-based compression

- MPEG-1 Audio Layer 3 – MP3

MPEG-1 Elementary Stream -- Video

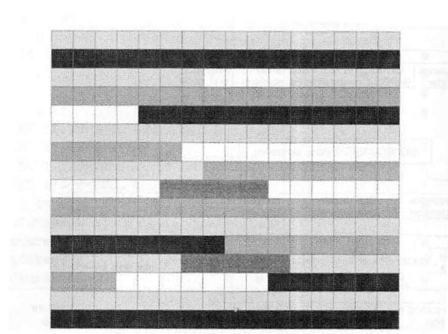
- Elementary Stream



E 6886 Topics in Signal Processing: Multimedia Security Systems

Intraframes – I frames

- Similar to JPEG compression
- Specify Slices

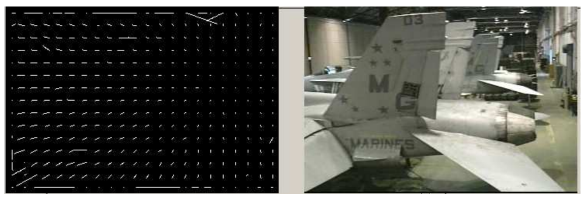


 © 2006 Ching-Yung Lin, Dept. of Electrical Engineering, Columbia Univ.

Motion Estimation -- I

- Motion Estimation Examples

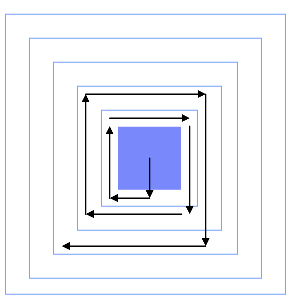
MV estimation



Zoom In

Motion Estimation -- II

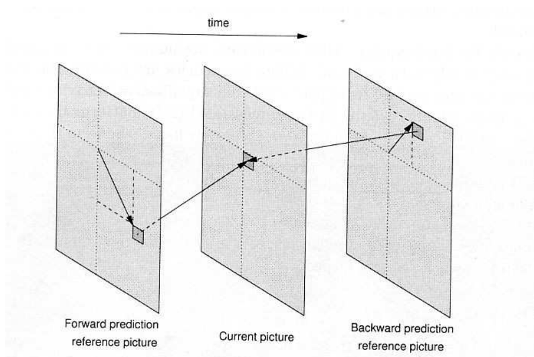
- Motion Estimation Example:
  - Move the original block and compare the distance with it's the corresponding blocks in the target frame.
  - Searching from nearest neighbors to outer locations. If the error distance between the original block and the shifted block in the target frame < a threshold, then quite searching
  - Search can be based on the motion vector of spatially adjacent blocks and previous (time = t – 1) motion estimation.



→ : greedy search pattern

Predicted Interframes – P and B frames

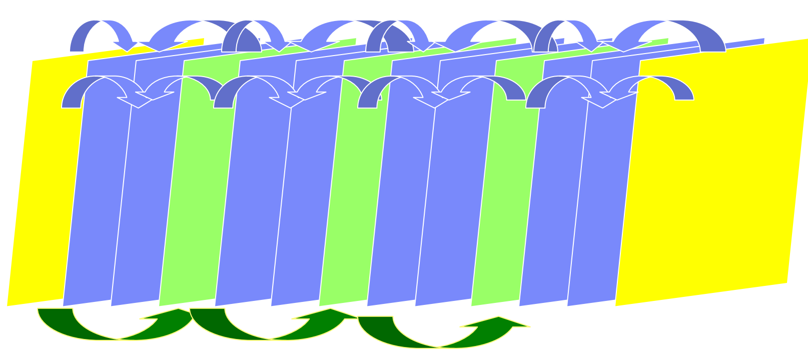
- Prediction:
  - P frames use only forward prediction
  - B frames use bi-directional prediction (average of both predictions)
- Coefficients encoding:
  - Residuals (current block – predicted block) are encoded using similar coding methods as in intraframe encoding
- Intrablocks:
  - P frames allow intrablock coding
  - B frames does not allow intrablock coding



Orders of I/P/B Frames

- Display Order:

I B B P B B P B B P B B I



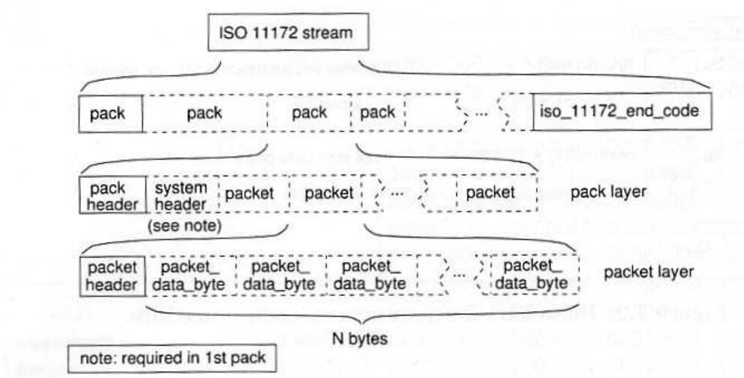
- Coding Order / Streaming Order:

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

  - Buffering is essential in the decoder

MPEG-1 System Stream

- System Stream: Audio + Video

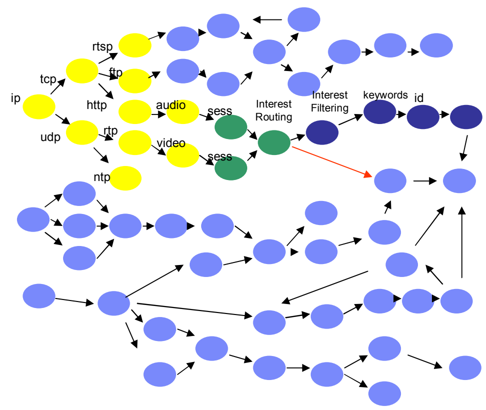


MPEG-2 v.s. MPEG-1

- MPEG-2 targets HDTV, film, video streaming applications.
- MPEG-2 has five profiles: Simple, Main, SNR Scalable, Spatially Scalable, and High
- Main Profile is similar to MPEG-1
- MPEG-2 system streams include all the features of MPEG-1 (with enhancements that allow it to operate in more error-prone environments.)
- MPEG-2 audio is backwards compatible with MPEG-1.

Video Streaming and Transcoding Demo

- Open Software: VideoLAN (http://www.videolan.org)



Advanced Video Coding (AVC) Standard: H.264/ MPEG-4 Part 10

- Aims to provide good video quality at bit rates that are substantially lower (e.g., half or less) than previous standards (MPEG-2, MPEG-4, H.263, etc.)
- No so much increase in complesity
- Main differences:
  - Allow both 4x4 and 8x8 transforms
  - Encoder-specified perceptual-based quantization
  - Efficient inter-picture lossless coding
  - Support additional color spaces and residual color transform
  - Motion compensation:
    - Multi-picture
    - Variable block-size
    - Quarter-pixel precision
    - Weighted prediction
- Applications:
  - HD-DVD format (Toshiba, NEC, Sanyo) (New Line Cinema, Paramount, Universal, Warner Brothers)
  - Blue-Ray Disc format (Sony, Matsushita, Samsung) (Sony, MGM, 20th Century Fox, Walt Disney)

References

- [JPEG] G. K. Wallace, The JPEG Still Picture Compression Standard, Communications of the ACM, 1991.
- [MPEG-1] J. Mitchell, W. B. Pennebaker, C. Fogg and D. J. LeGall, MPEG Video Compression Standard, Chapman and Hall, 1996
- [MPEG-2] B. Haskell, A. Puri and A. Nettravali, Digital Video: An Introduction to MPEG-2, Chapman and Hall, 1997
- [Motion Estimation] A. M. Tekalp, Digital Video Processing, Prentice Hall, 1995
- [AVC/H.264] I. Richardson, H.264 and MPEG-4 Video Compressions: Video Coding for Next Generation Multimedia, 2003

