Advanced Computer Networks

Video Basics

Jianping Pan
Spring 2017

Video is a sequence of images

- Recorded/displayed at a certain rate
- Types of video signals
  - component video
    - separate RGB signals; e.g., VGA CRT
  - composite video
    - luminance and chrominance in one signal carrier
  - S-video
    - 1 luminance and 1 composite chrominance signal
  - VGA, DVI, HDMI, etc

Video

- Image
  - picture resolution: e.g., 640x480
  - pixel depth: e.g., 8-bit
- Video
  - frame rate > flicker-free rate
    - movie: 24 frames/second
    - TV: 25 or 30 frames/second
    - VGA CRT: e.g., 50Hz

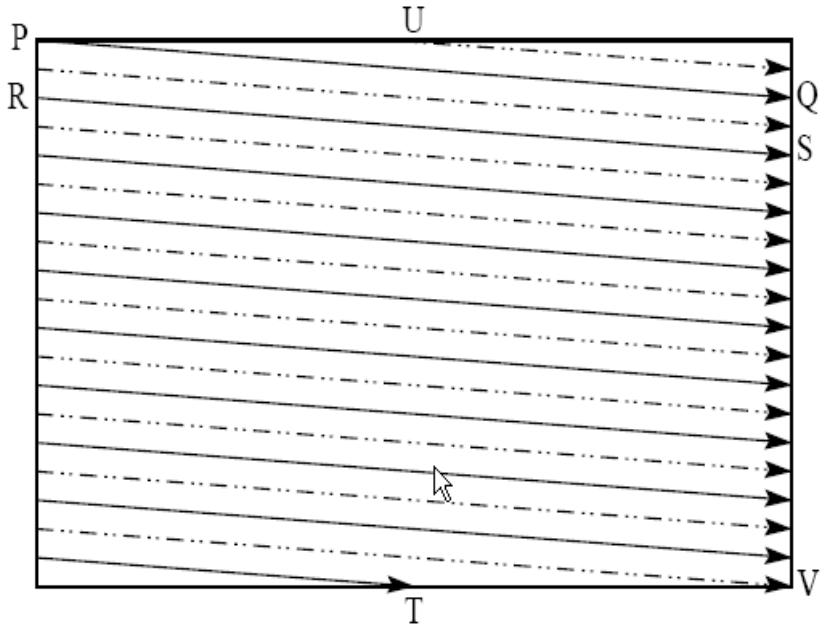
newer: 120fps

We see most video on

- Television
  - PAL (Phase Alternating Line)
    - 625 lines interlaced (576 visible)
    - 25 frames/second
    - aspect ratio 4:3
    - YUV
  - NTSC (National TV Standards Committee)
    - 525 lines interlaced (480 visible)
    - 30 frames/second (29.97 to be exact), 4:3, YIQ

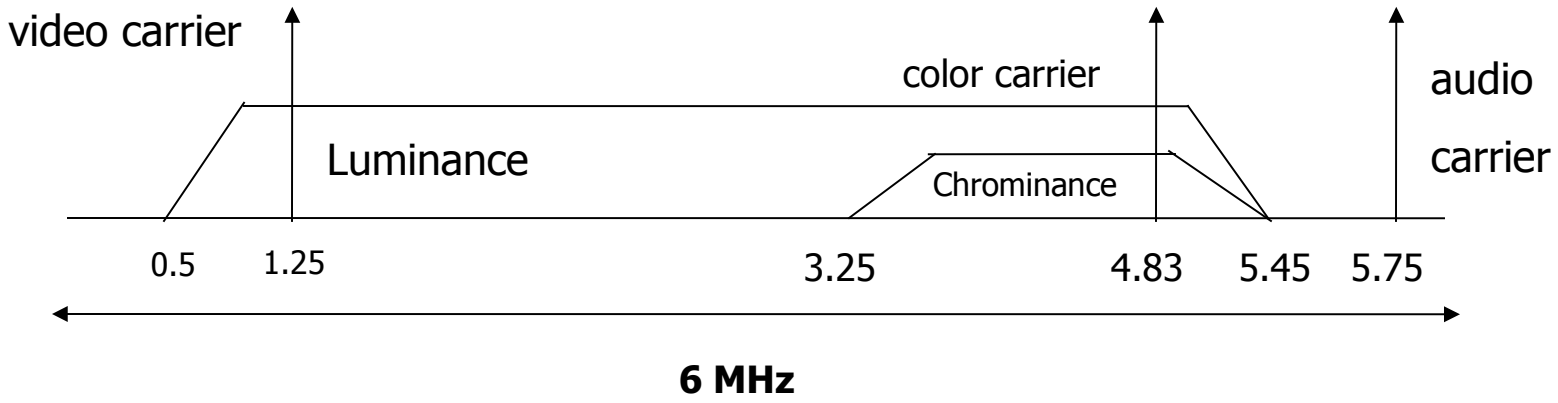
Interlaced vs progressive

- Interlaced
  - odd line: P => Q
  - Q => R (H retrace)
  - R => S
  - …
  - T => U (V retrace)
  - even line: dash-dot
- Progressive



TV broadcasting

- NTSC (6MHz channel)
  - lower band: guard; upper band: audio (FM)
  - Y: 4.2MHz
  - I: 1.6MHz; Q: 0.6MHz



now go digital white space?

Digital video

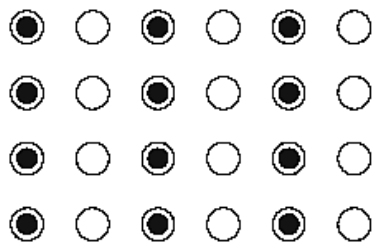
Standards for Video

	HDTV	CCIR 601 NTSC	CCIR 601 PAL	CIF	QCIF
Luminance Resolution	1920 x 1080	720 x 486	720 x 576	352 x 288	176 x 144
Chrominance Resolution	960 x 540	360 x 486	360 x 576	176 x 144	88 x 72
Color Subsampling	4:2:2	4:2:2	4:2:2	4:2:0	4:2:0
Fields/sec	60	60	50	30	30
Aspect Ratio	16:9	4:3	4:3	4:3	4:3
Interlacing	Yes	Yes	Yes	No	No

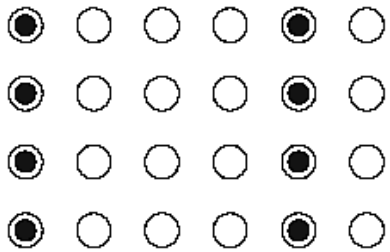
CCIR – Consultative Committee for International Radio
CIF – Common Intermediate Format (approximately VHS quality)

Chroma subsampling

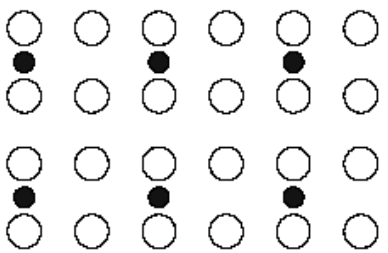
- 4:4:4: no subsampling
- 4:2:2, 4:1:1: chroma as 1/2 or 1/4 luma
- 4:2:0: vertical subsampling as well



4:2:2



4:1:1



4:2:0

Chroma subsampling examples

- Common Intermediate Format (CIF)
  - 4:2:0
  - Y: 352 x 288; U and V: 176 x 144
- Quarter CIF (QCIF): 176x144; 4:2:0

Y: 176 x 144

U: 88 x 72

V: 88 x 72

HDTV

- High Definition TV: better video/audio

# of Active Pixels per line	# of Active Lines	Aspect Ratio	Picture Rate
1,920	1,080	16:9	60I 30P 24P
1,280	720	16:9	60P 30P 24P
704	480	16:9 & 4:3	60I 60P 30P 24P
640	480	4:3	60I 60P 30P 24P

TV resolution evolution

- LDTV: low definition
  - 240i60, 288i50
- SDTV: standard definition
  - 480i60, 480p30, 576i50, 576p25
- EDTV: enhanced definition
  - 480p60, 576p50, 720i50/60, 720p24/25/30
- HDTV: high definition
  - 720p50/60, 1080p24/25/30, 1080i50, 1080i60

4KTV, 8KTV?

Temporal redundancy

- Video is a sequence of images
  - e.g., motion JPEG: M-JPEG
- Correlation between consecutive images
  - "difference" due to object or camera motion

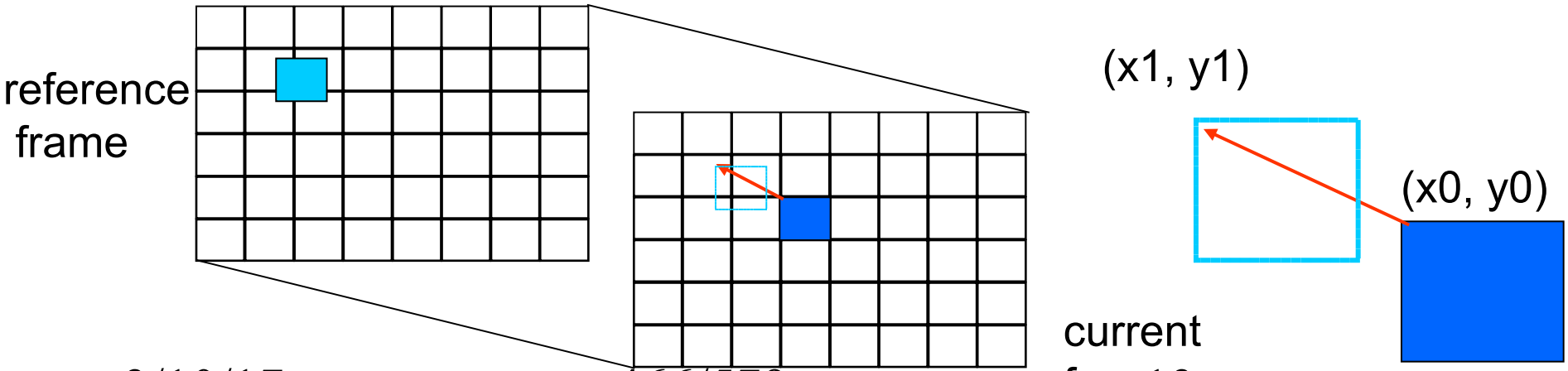
Frame i

Frame i+1

Direct Difference

Motion estimation

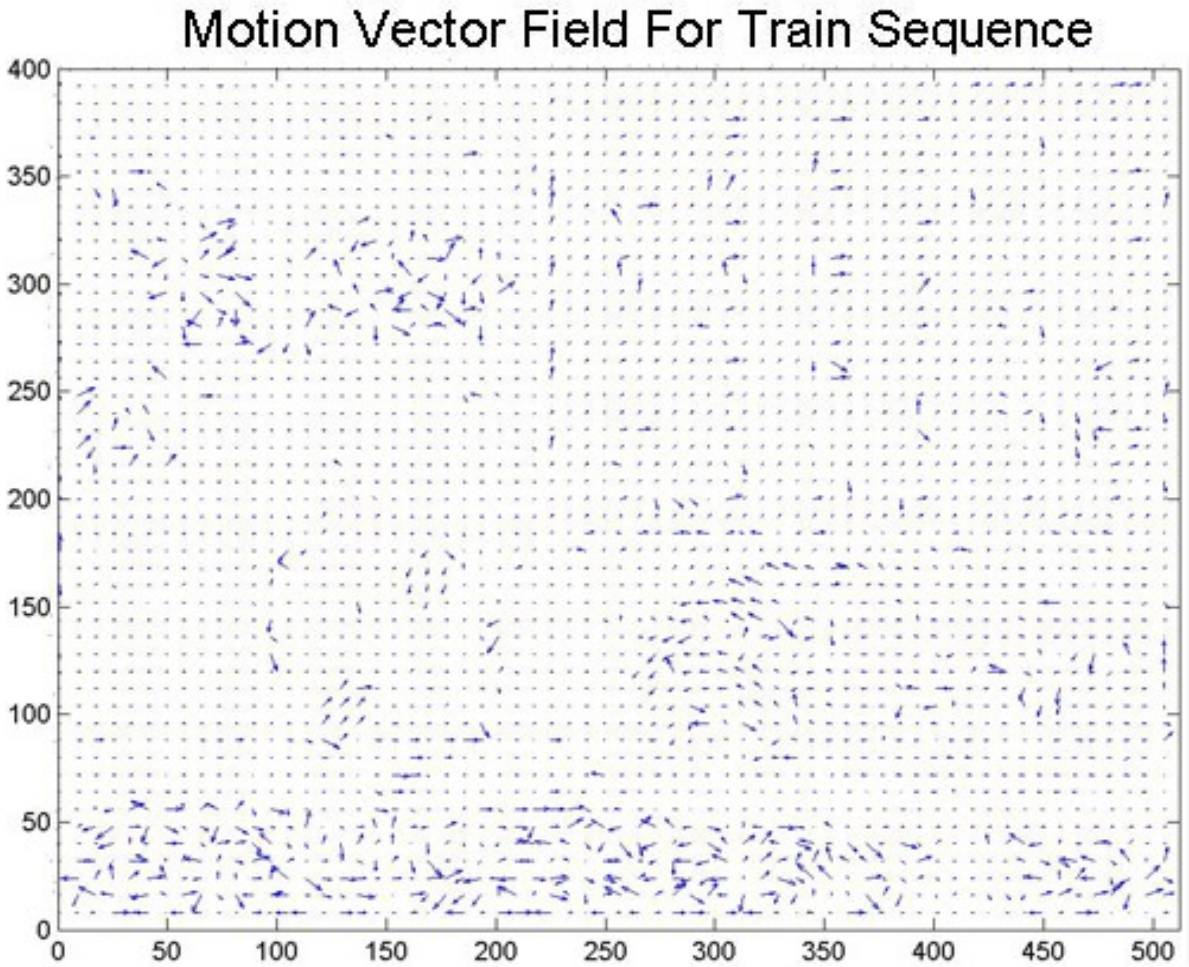
- Macro-block: 16x16 pixels
  - find a *similar* macro-block in the reference frame
  - record the motion "vector": (dx,dy)=(x1-x0,y1-y0)
  - encode the "difference" between two macro-blocks



Motion vector example

Anchor Frame

Target Frame



Macro-block similarity

- Similarity measures
  - mean square error (MSE)
  - mean absolute distance (MAD)

$$MAD(i,j) = \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} |C(x+k, y+l) - R(x+i+k, y+j+l)|$$

N – size of the macroblock,

k and l – indices for pixels in the macroblock,

i and j – horizontal and vertical displacements,

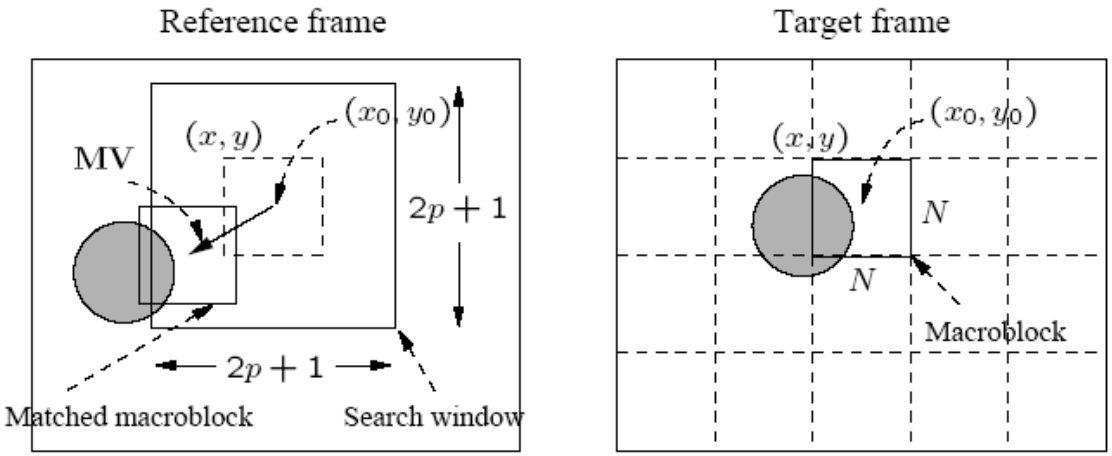
$C(x+k, y+l)$ – pixels in macroblock in Target frame,

$R(x+i+k, y+j+l)$ – pixels in macroblock in Reference frame.

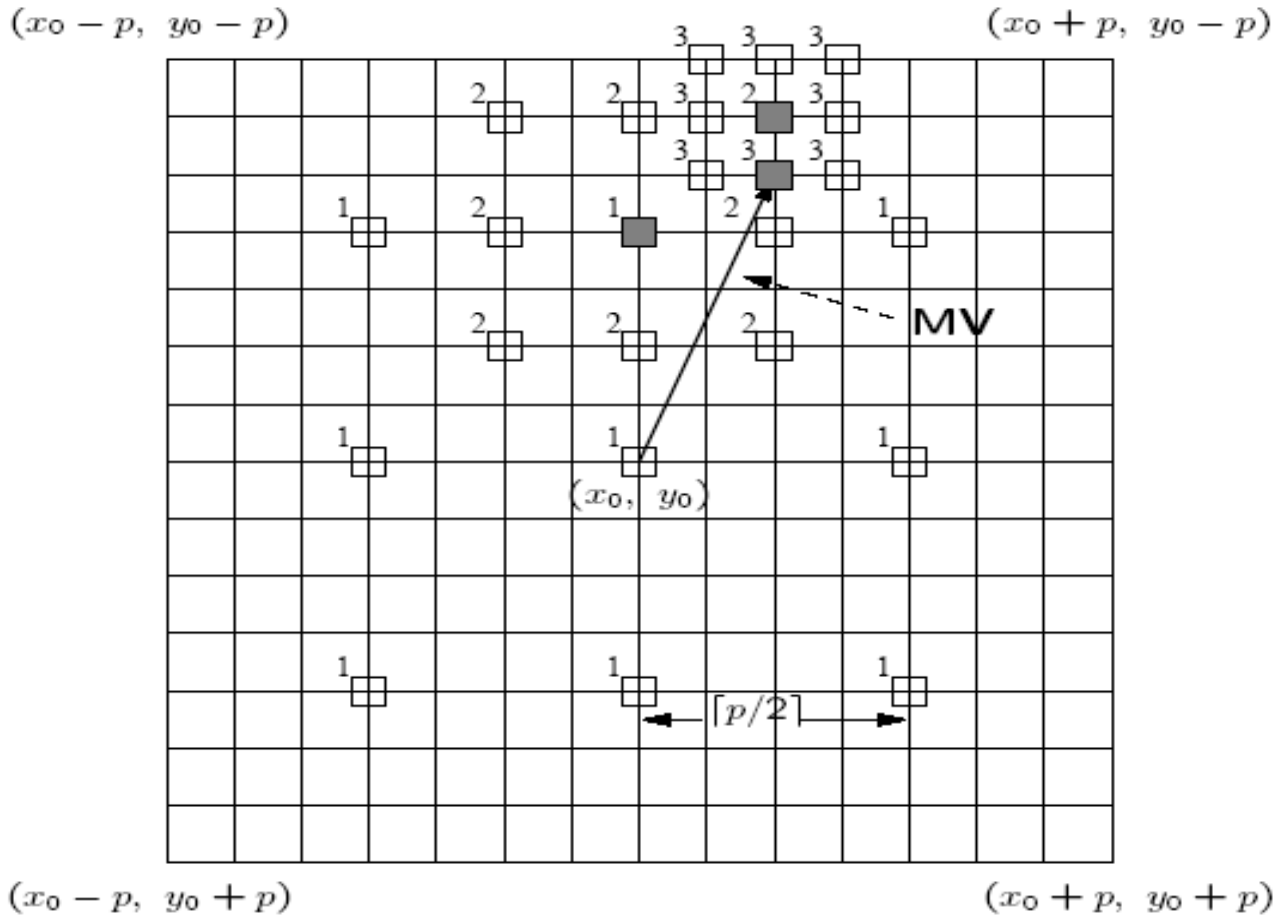
slower encoder?

Search window

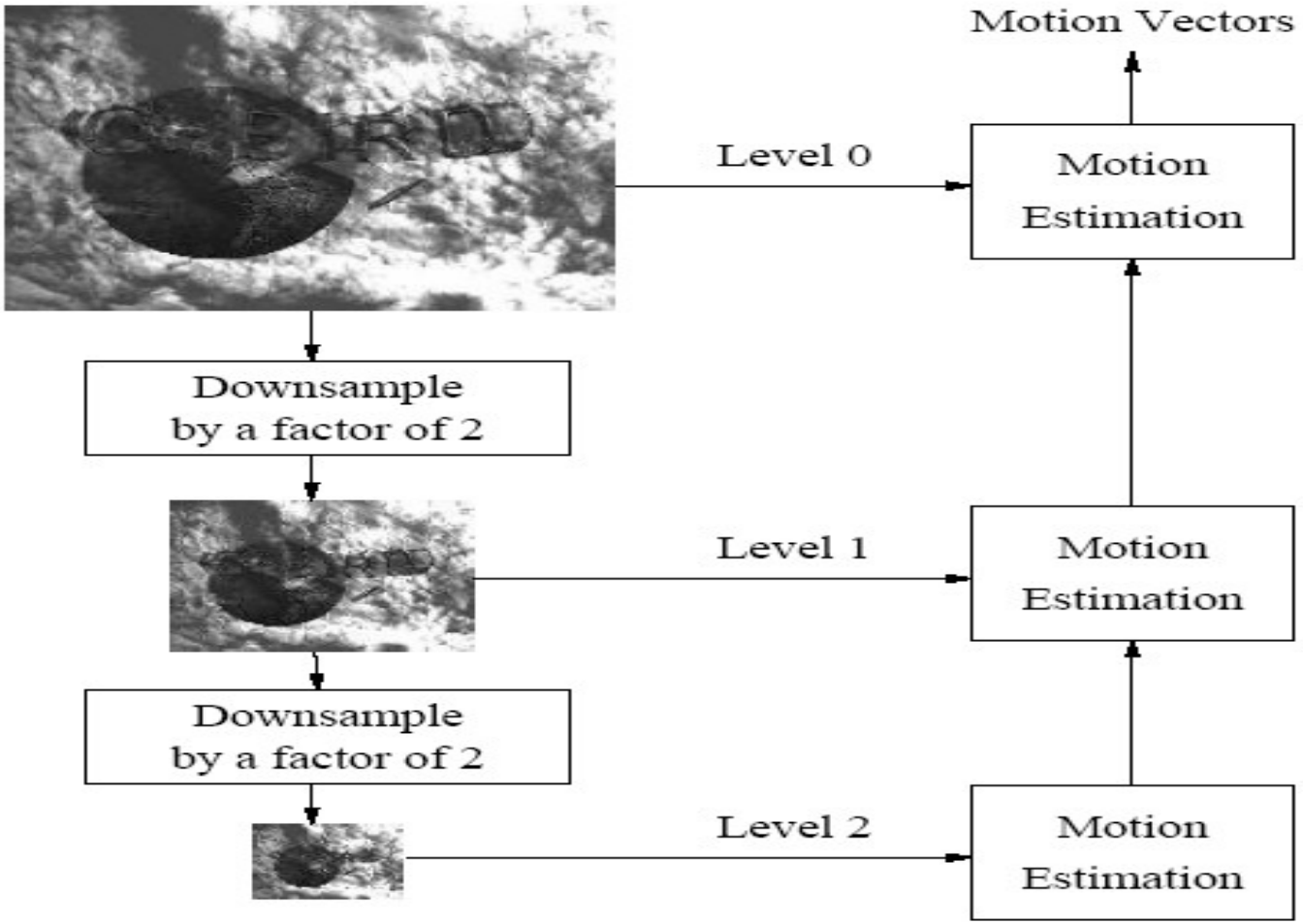
- Rectangle: x: [x0-p, x0+p]; y: [y0-p,y0+p]
- $(2p+1)^2$ all possible reference macro-blocks
  - need *better* search algorithms!



2-D Log motion search

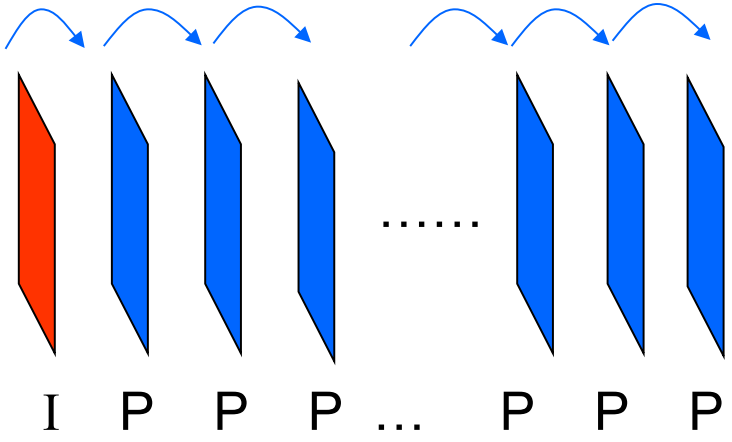


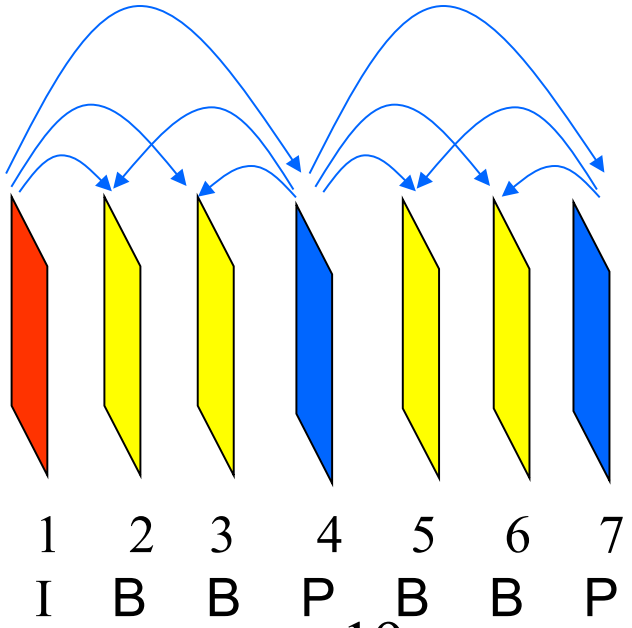
Hierarchical motion search



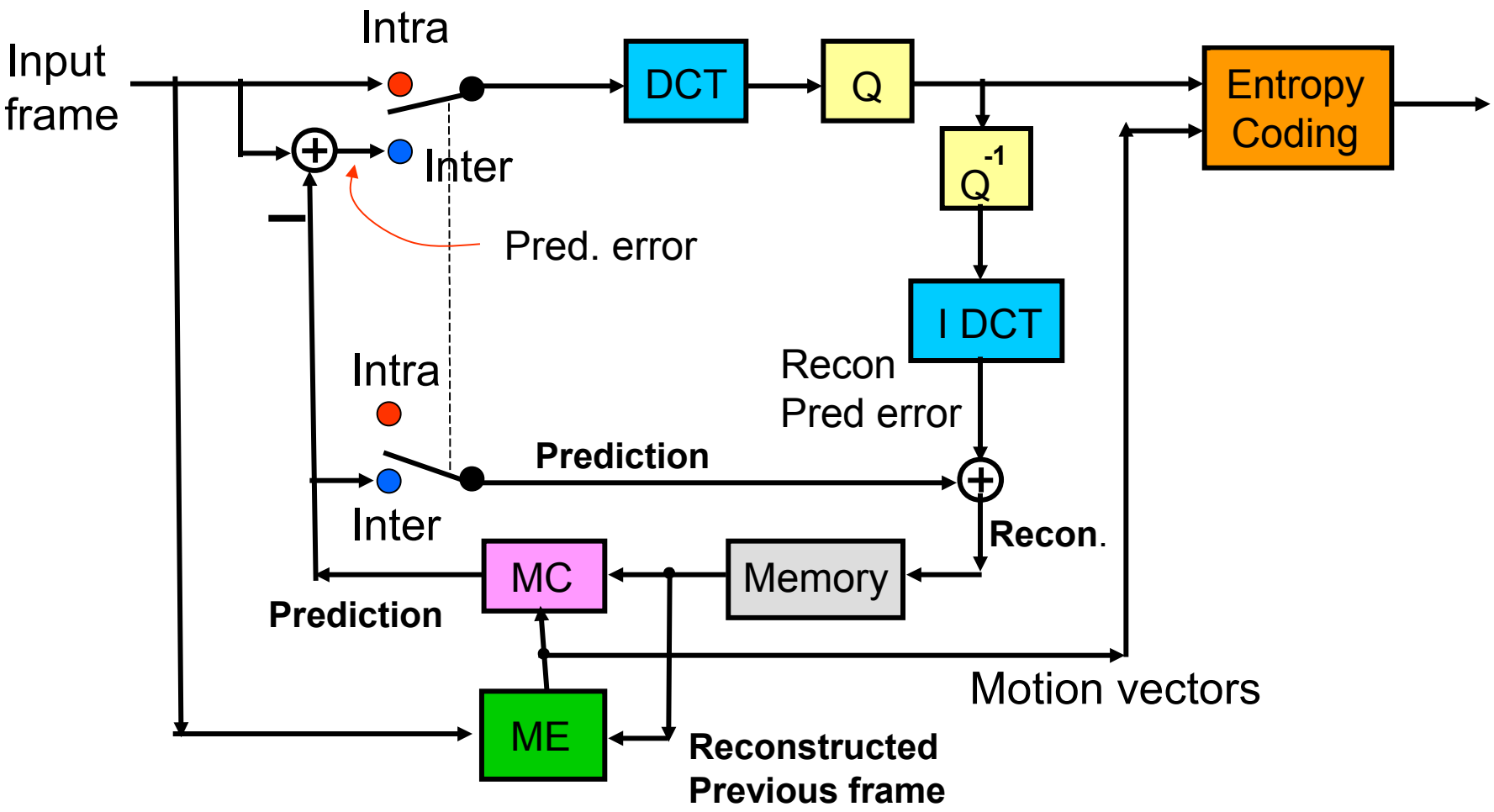
Group of pictures

- B: bidirectionally interpolated frame
- P: predicted frame
- I: intra-coded frame



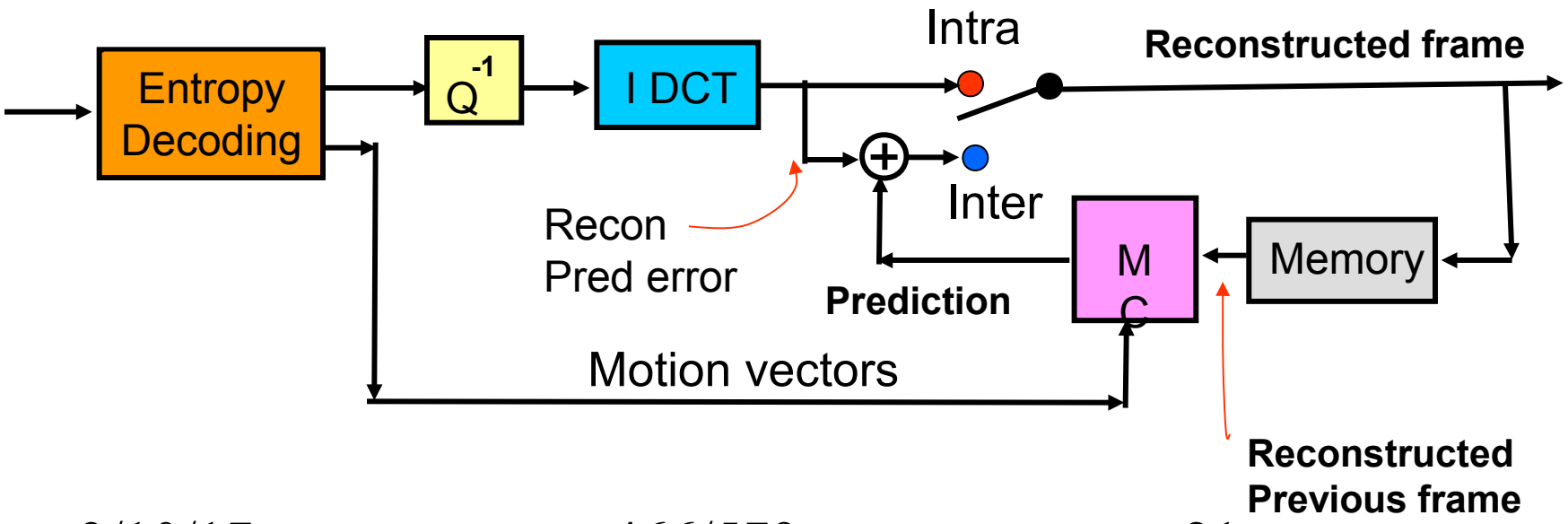


Video encoder



Video decoder

- Decoder is simpler than encoder
  - usually only the decoder is standardized
  - allow innovations at encoders

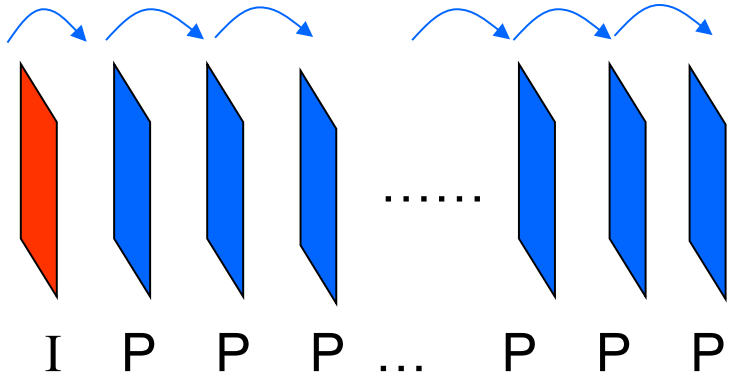


H.261

- H.261: *p**64Kbps (*p*: 1~30)
  - ITU-T recommendation (1990)
  - real-time video telephony over ISDN (2B+D)
    - end-to-end delay less than 150ms
  - QCIF (required): 176x144, 4:2:0, ~30fps,3 GOB
  - CIF (optional): 352x288, 4:2:0, ~30fps, 12GOB
  - GOB: group of 3x11 macro-blocks
    - 1 macro-block: 4 Y block, 1 Cr block, 1 Cb block
    - 1 block: 8x8 pixel (e.g., in luminance)

H.261: more

- I-frame (JPEG-like)
  - RGB=>YUV, 8x8 blocks
  - DCT
  - Scalar quantization
  - ZigZag scanning, DC/AC encoding, entropy encoding
- P-frame
  - search window p=15
  - pixel precision



H.263

- H.263: initially < 64Kbps; later higher bps
  - ITU-T Rec (1995); v2(1998); v3 (2000)
- More video formats
  - sub-QCIF, QCIF, CIF, 4CIF, 16CIF
- More motion estimation techniques
  - half-pixel precision
  - modes: unrestricted motion vector, arithmetic coding, advanced prediction, PB-frames, etc

MPEG

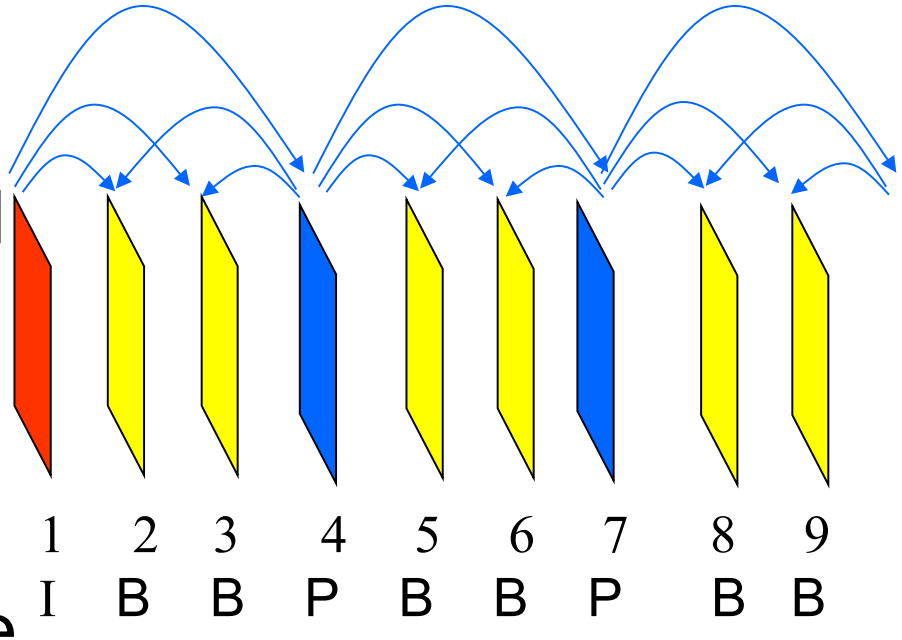
- Motion Picture Experts Group
  - MPEG-1: VCD (VCR-quality)
  - MPEG-2: DVD & HDTV
  - MPEG-3: aborted due to MPEG-2
  - MPEG-4: content-based
  - (future compression standards)
  - MPEG-7: meta-data
  - MPEG-21: DRM (21st century)

MPEG-1

- MPEG-1 (1991): VCD (VCR+CD quality)
  - 352x240, 1.2Mbps video CBR, 256Kbps audio
  - progressive scan only (1x CD-ROM)
- MPEG-1 video compression
  - similar to H.261, with a few differences
    - more formats, flexible *slice*s, quantization table
  - I-frame: JPEG-like compression
  - P-frame: prediction-based; B-frame

MPEG-1: more

- Bi-directional search
  - search both previous and next frames for *similar* macro-blocks
- MPEG-1 GOP
  - I-frame, P-frame, B-frame
    - display order: IBBPBBPBBPBBPBBI (M=3, N=15)
    - coding order: IPBBPBBPBBPBBIBB; timestamps
  - D-frame: for search through the video, DC only



MPEG-2

- MPEG-2 (1994): DVD, HDTV, etc
  - also adopted as ITU-T H.262
  - many video formats and data rates; better audio
    - profiles: simple (4:2:0, I/P), main (+B), SNR (+variable quality), spatial (+variable resolution), high (+4:2:2)
    - levels: low (352x288), main (720x576), high 1440 (1440x1152), high (1920x1152)
  - support interlaced video (broadcasting!)

LEVELS and PROFILES	Simple Profile	Main Profile	SNR Scalable Profile	Spatial Scalable Profile	High Profile
	No B-frames	B-frames	B-frames	B-frames	B-frames
	4:2:0	4:2:0	4:2:0	4:2:0	4:2:0 or 4:2:2
	Not Scalable	Not Scalable	SNR Scalable	SNR Scalable or Spatial Scalable	SNR Scalable or Spatial Scalable
High Level 1920 pixels/ line 1152 lines		≤ 80 Mbit/ s			≤ 100 Mbit/ s
High-1440 Level 1440 pixels/ line 1152 lines		≤ 60 Mbit/ s		≤ 60 Mbit/s	≤ 80 Mbit/s
Main Level 720 pixels/ line 576 lines	≤ 15 Mbit/ s	≤ 15 Mbit/ s	≤ 15 Mbit/ s		≤ 20 Mbit/s
Low Level 352 pixels/ line 288 lines		≤ 4 Mbit/s	≤ 4 Mbit/s		

MPEG-2 profiles and levels

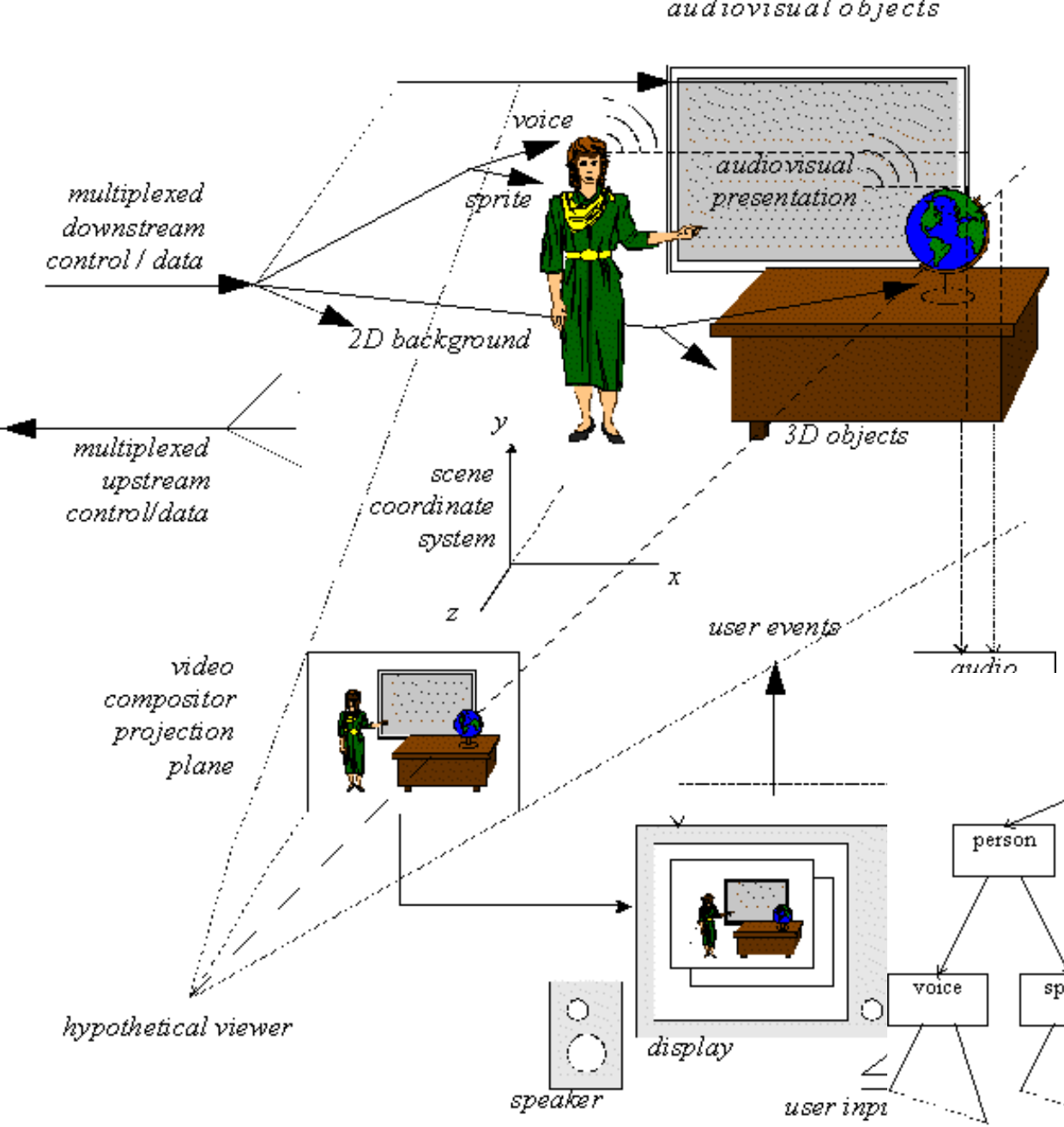
MPEG-2 scalability

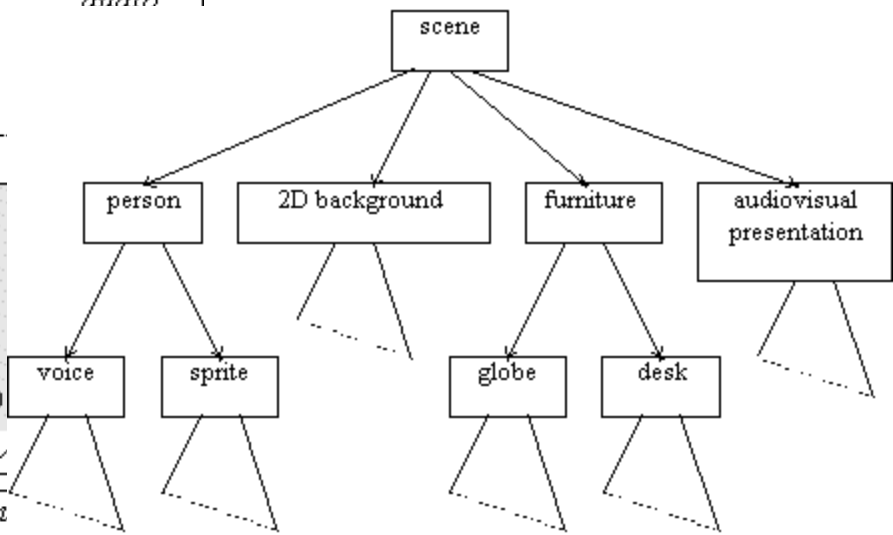
- Layered encoding
  - base layer: independent for basic quality
  - enhancement layer: dependent on the base layer
- E.g., SNR scalability
  - base: low SQNR (coarse quantization)
  - enhance: high SQNR (fine Q on actual-base)
- E.g., spatial scalability
  - base: low resolution; enhance: high resolution

MPEG-4

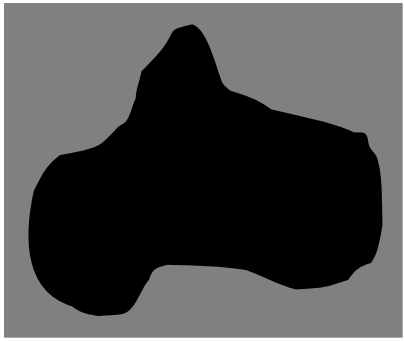
- MPEG-4 (1999): content-based, object-oriented
  - based on H.263, initially for low bit-rate apps
  - video sequence: a collection of media *object*s
    - objects: still image, moving object, audio, etc
    - how to decompose is NOT specified (encoder)
  - VOP: video object plane
    - GOV: I-VOP, P-VOP, B-VOP
    - VOP is divided into many macro-blocks
  - motion estimation: bounding box; padding

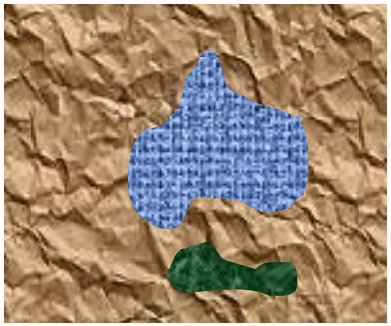
MPEG-4: object oriented

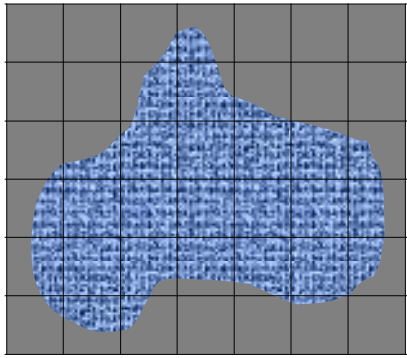




Object coding

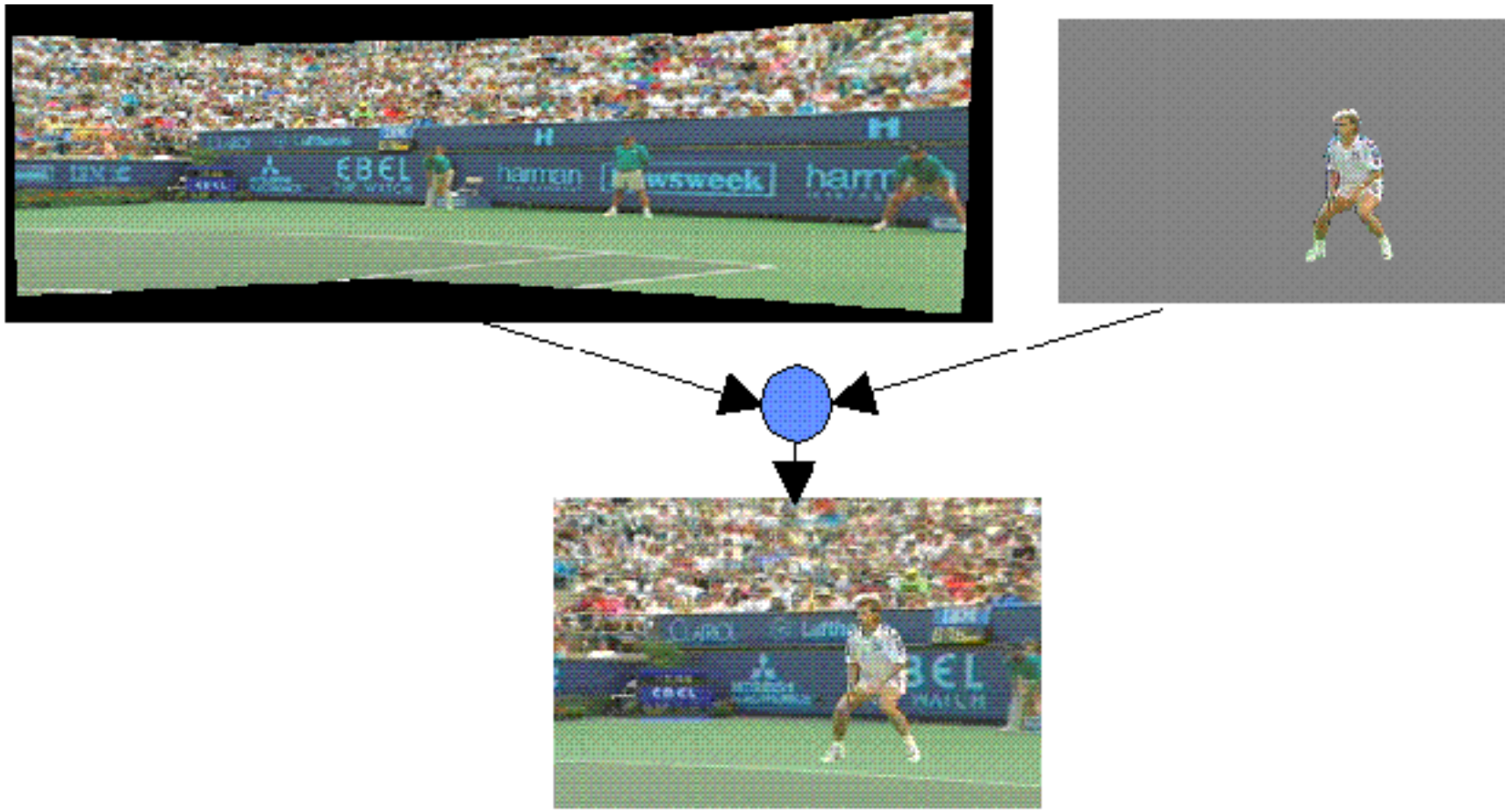






- Texture coding
  - DCT-based
  - SA-DCT: shape adaptive
- Shape coding
  - binary shape; grayscale (transparency) shape
- Static texture coding
  - wavelet-based (good for scaling)
- 2-D and 3-D mesh coding

Sprite coding



MPEG-4: more

- Fine gain scalability
  - spatial scalability
  - temporal scalability
  - quality scalability
- MPEG-4 audio
  - general audio (2~64Kbps)
  - speech (2~4Kbps: HVXC; 4~24Kbps: CELP)
  - synthesized (e.g., MIDI, TTS)

H.264

- H.264 (2003)
  - also as MPEG-4 AVC (advance video coding)
  - initially: low data rate for high picture quality
  - now a wide variety of bit-rates, applications, systems
  - enhanced motion estimation and compensation
    - multi-picture, variable block-size, quarter-pixel precision, weighted prediction, etc
  - profiles: baseline, main, extended; 15 levels
  - fidelity range extension: high, 10, 4:2:2, 4:4:4

* H.265 (HEVC)? SVC,MVC, etc

This lecture

- Video representation
- Video compression
  - motion vector
    - how to find a similar macro-block
  - generic video encoder/decoder
- Video standards
  - H.263v2 (H.263+) and H.263v3 (H.263++)
  - MPEG/AVC, H.264, H.265