

CSc 461/561
Multimedia Systems
Video compression

Jianping Pan
Spring 2015

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

2/4/15

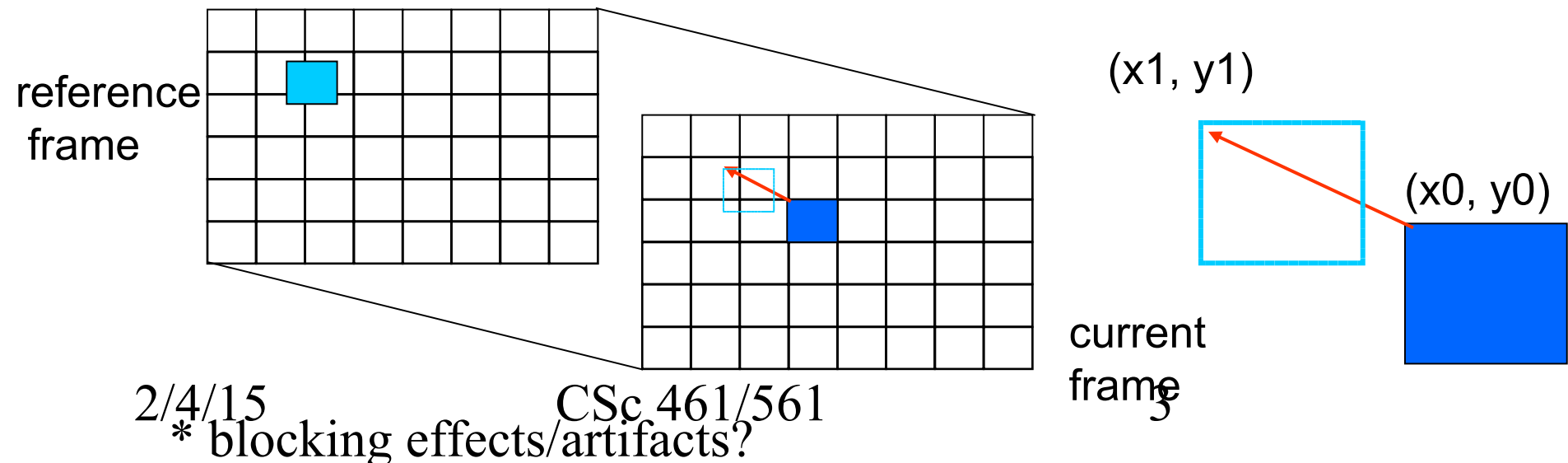
CSc 461/561

2

* refresh rate? motion objects vs pan-tilt-zoom (PTZ) cameras

Motion estimation

- Macro-block: 16x16 pixels
 - find a *similar* macro-block in the reference frame
 - record the motion “vector”: $(dx, dy) = (x_1 - x_0, y_1 - y_0)$
 - encode the “difference” between two macro-blocks



Motion vector example

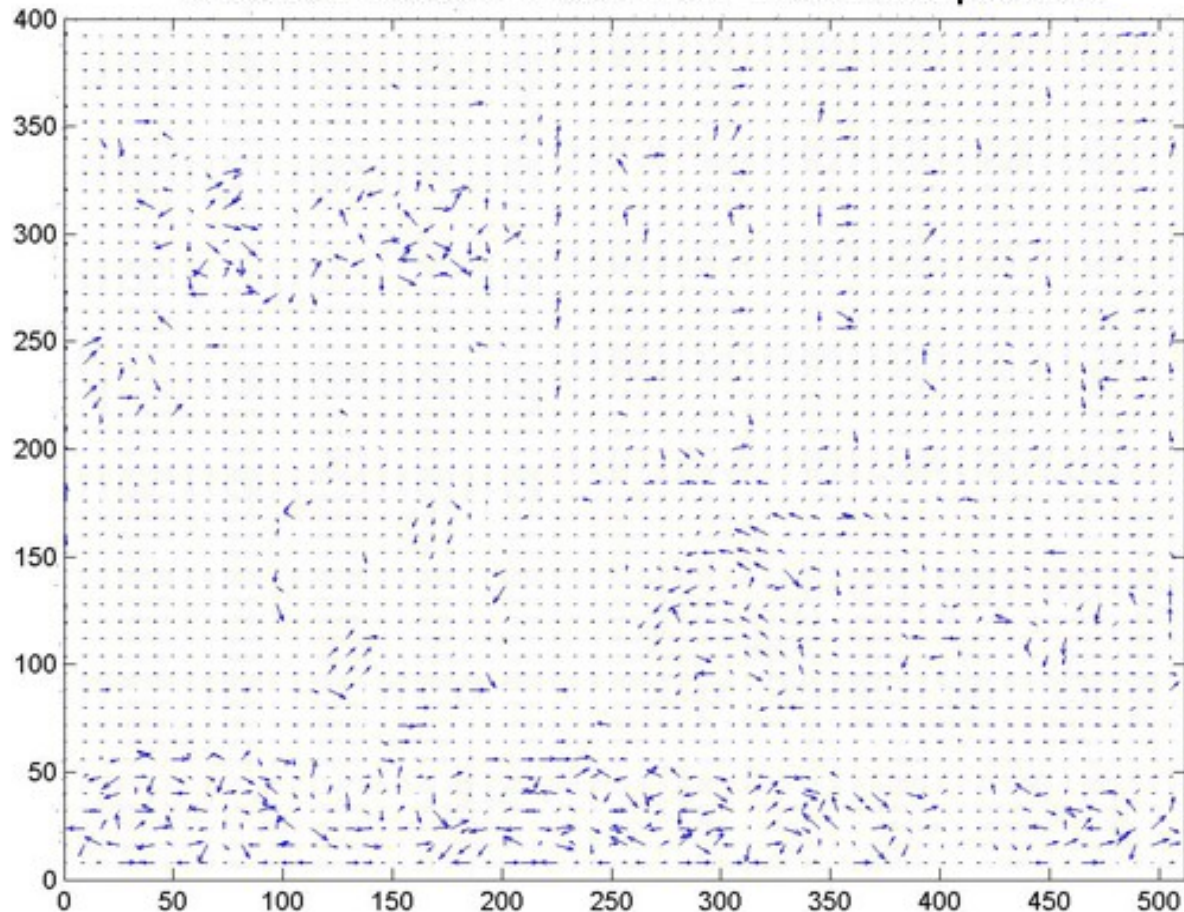
Anchor Frame



Target Frame



Motion Vector Field For Train Sequence



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,

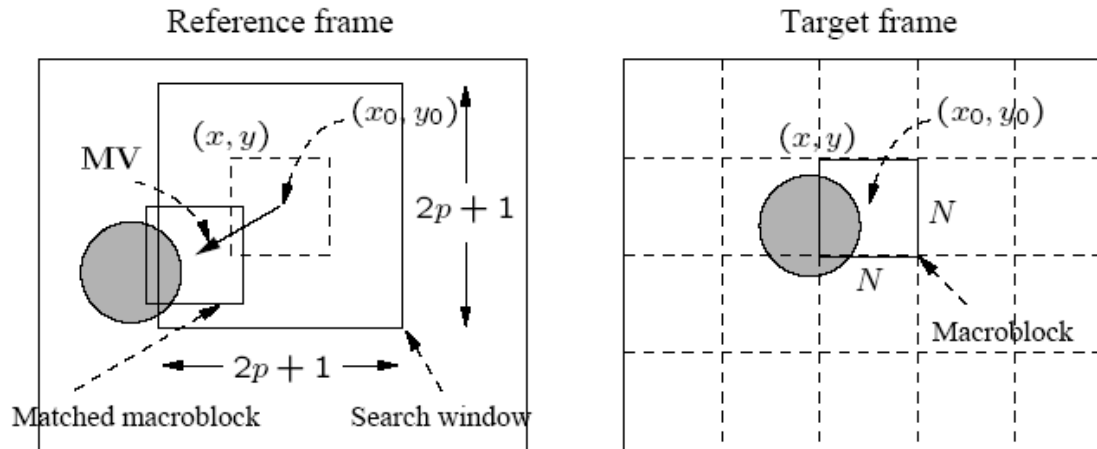
i and j – horizontal and vertical displacements,

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

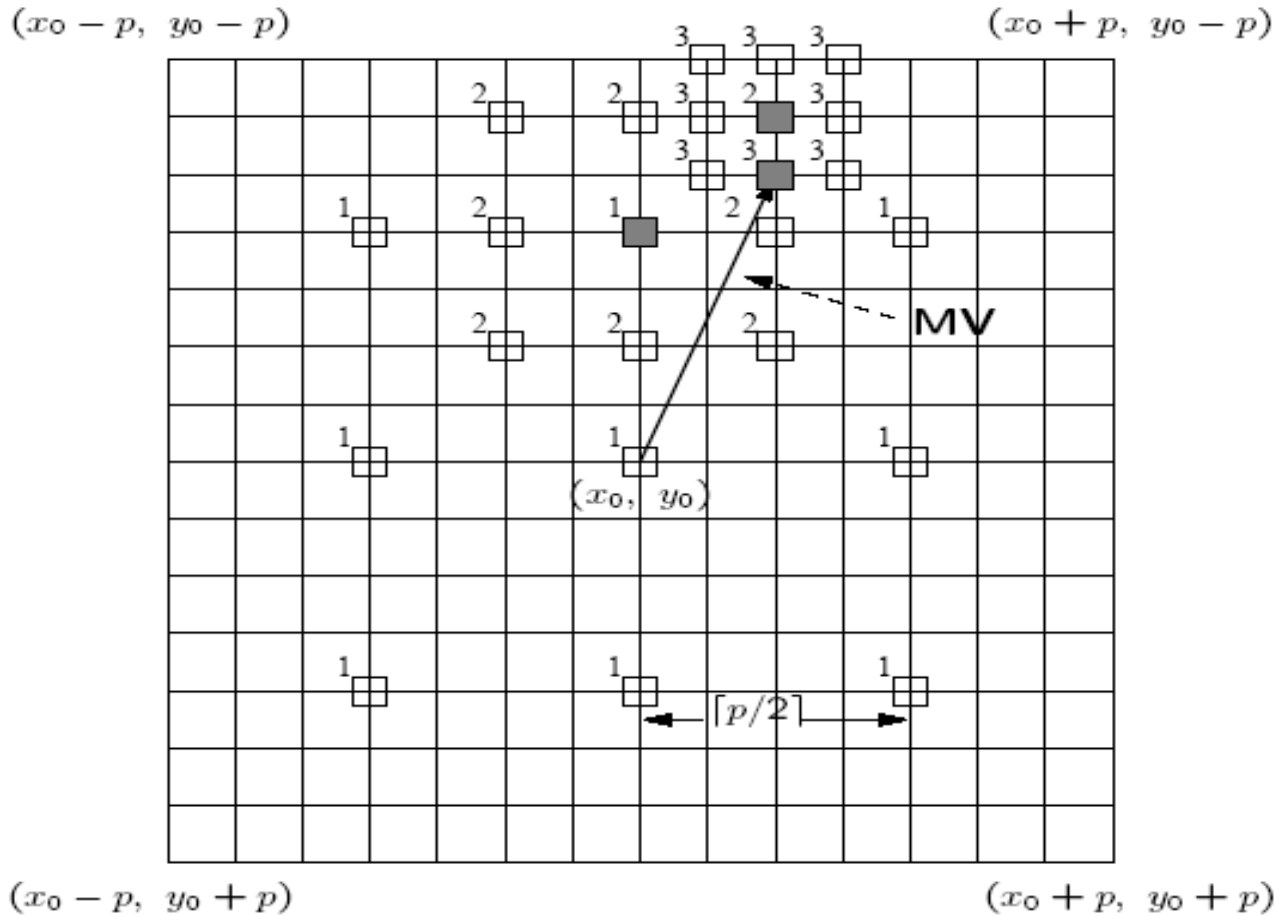
$R(x + i + k, y + j + l)$ – pixels in macroblock in Reference frame.

Search window

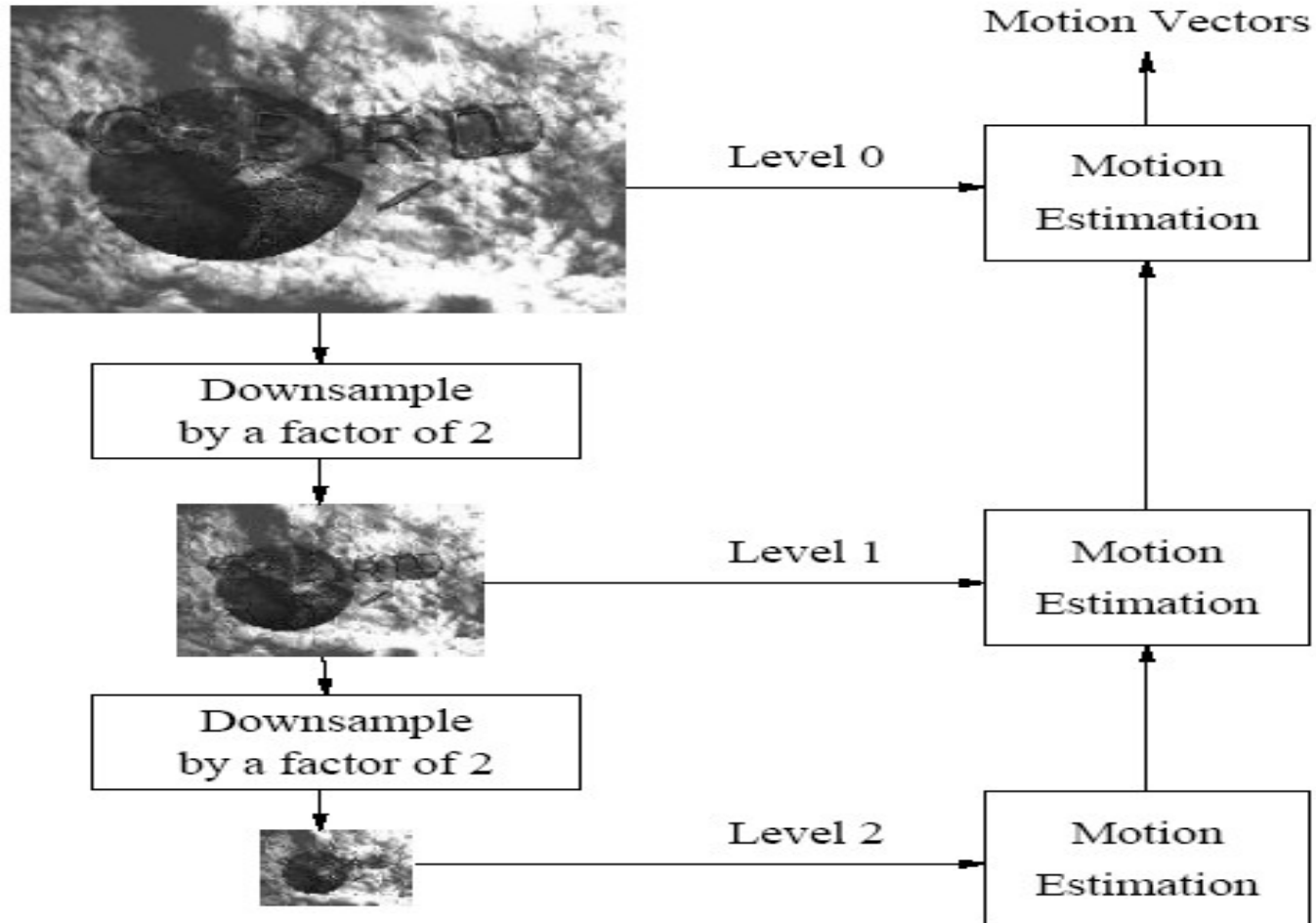
- Rectangle: $x: [x_0-p, x_0+p]$; $y: [y_0-p, y_0+p]$
- $(2p+1)^2$ all possible reference macro-blocks
 - need *better* search algorithms!



2-D Log motion search

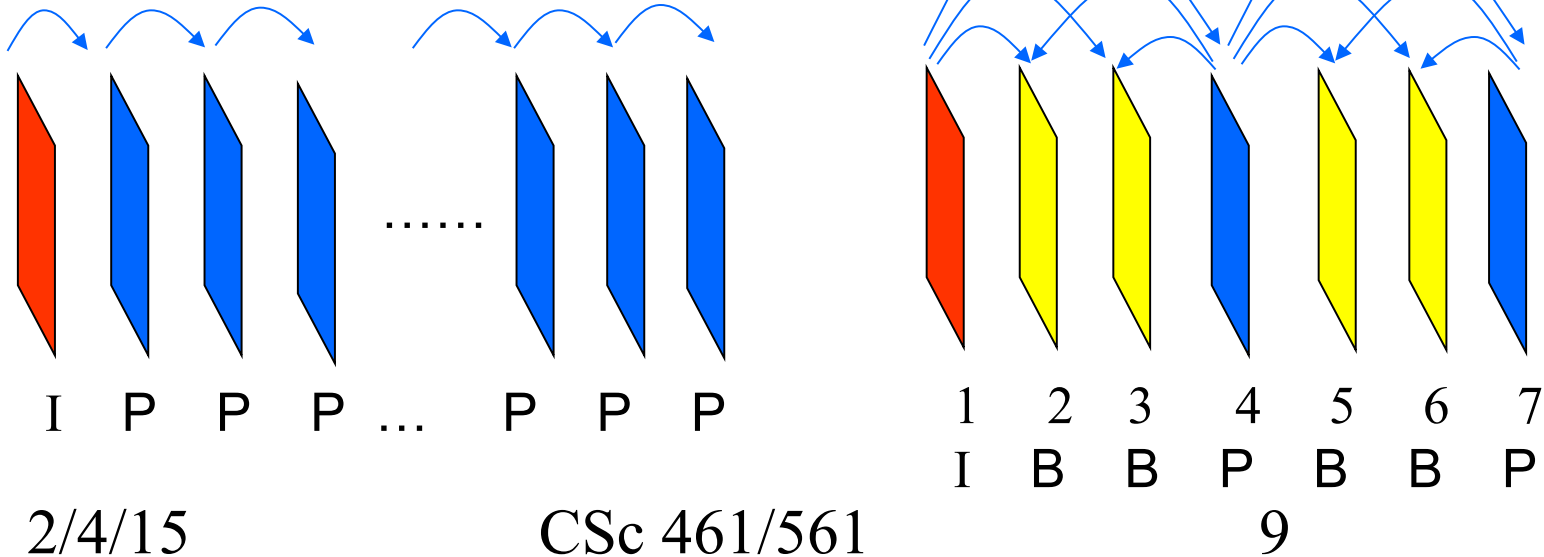


Hierarchical motion search

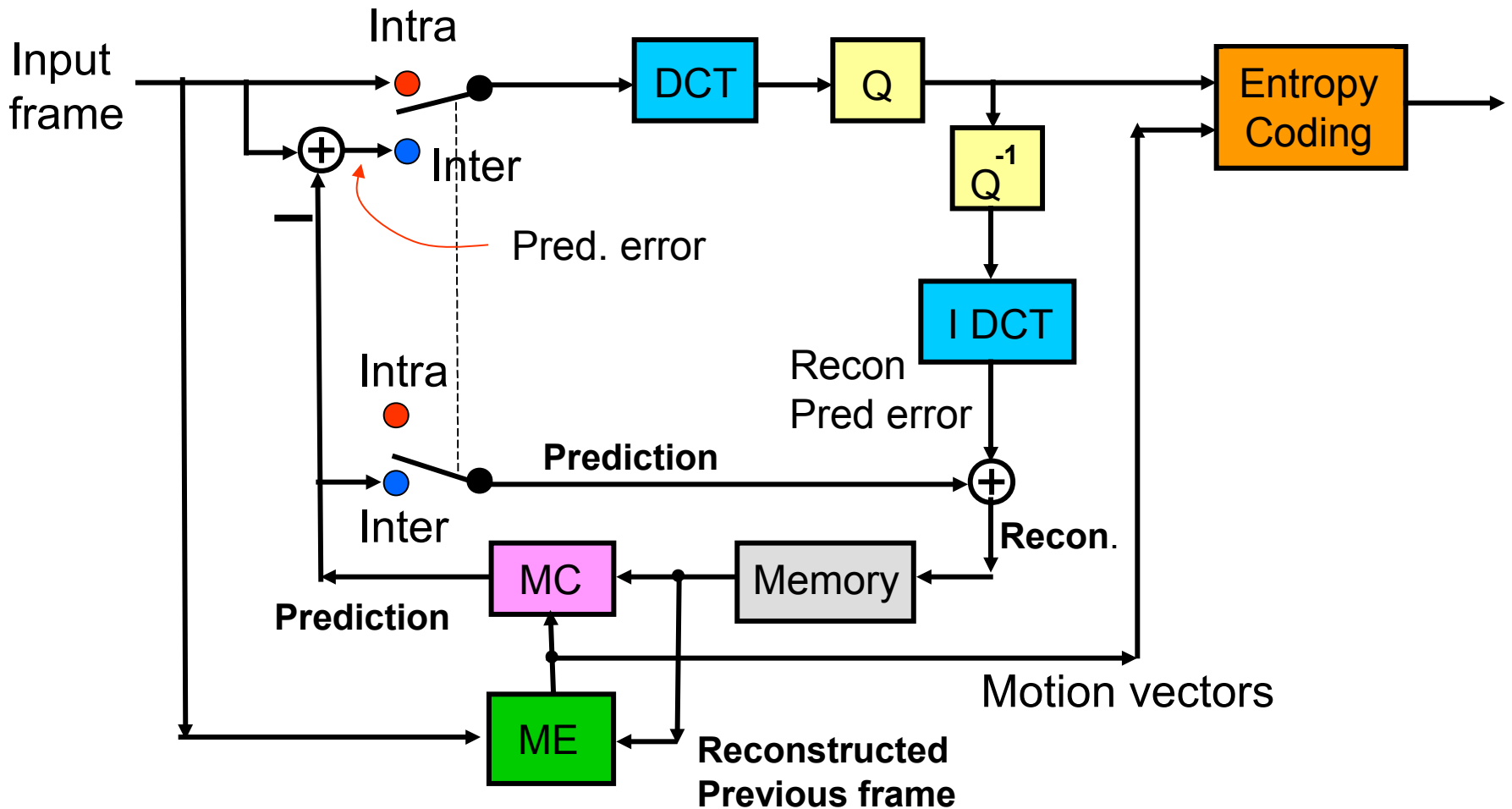


Group of pictures

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

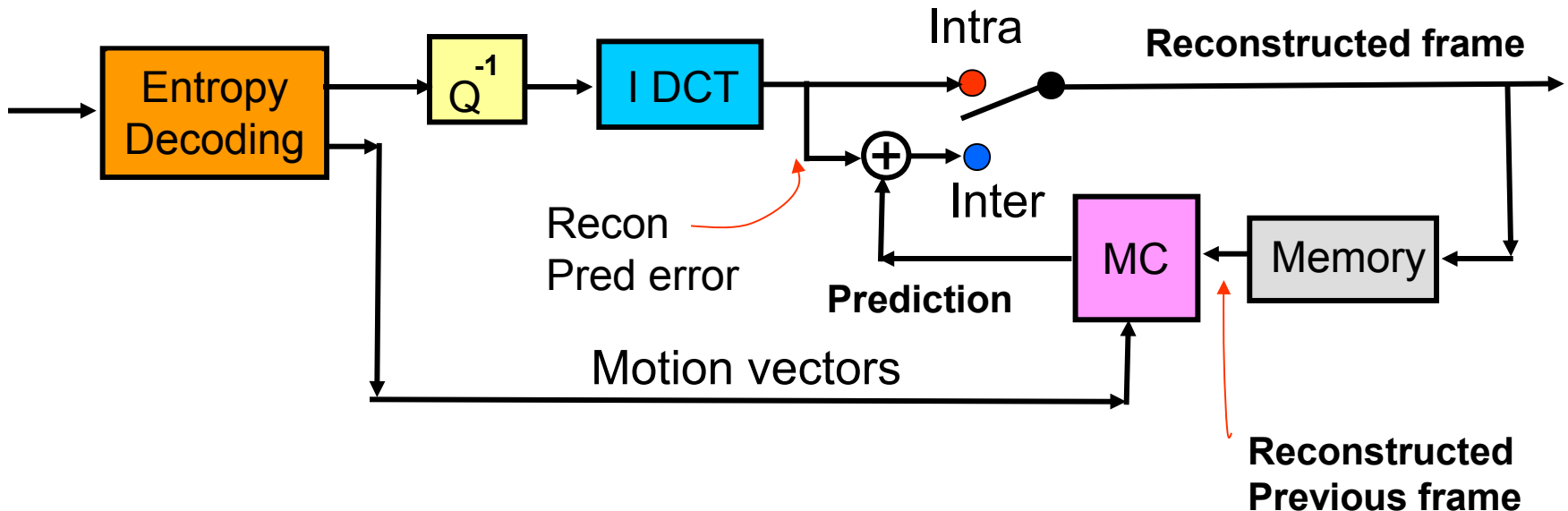


Video encoder



Video decoder

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



H.261

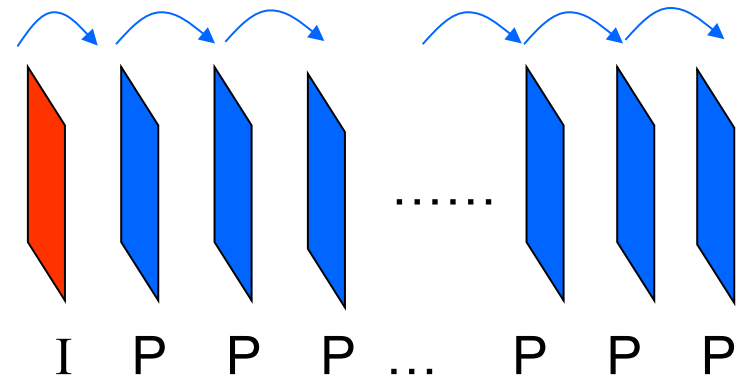
- H.261: $p * 64\text{Kbps}$ ($p: 1 \sim 30$)
 - ITU-T recommendation (1990)
 - real-time video telephony over ISDN (2B+D)
 - end-to-end delay less than 150ms
 - QCIF (required): 176x144, 4:2:0, ~30fps, 3 GOB
 - CIF (optional): 352x288, 4:2:0, ~30fps, 12GOB
 - GOB: group of 3x11 macro-blocks
 - 1 macro-block: 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 $< 64\text{Kbps}$; 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

This lecture

- Video compression
 - motion vector
 - how to find a similar macro-block
 - generic video encoder/decoder
 - examples: H.261/263
- Explore further
 - H.263v2 (H.263+) and H.263v3 (H.263++)

Next lecture

- Multimedia manipulation
 - video compression standards
[Ref: Li&Drew Chap 11-12]
 - MPEG-1/2/4 [11.2-3, 12.1-2]
 - H.264 [12.5]