Digital Image Processing (CSE/ECE 478)

Lecture # 22: Motion estimation and video compression

Avinash Sharma

Center for Visual Information Technology (CVIT),

IIIT Hyderabad

# Videos

- A sequence of still frames shown together

Videos

- Origin of motion picture takes us to popularly debated question of those times:

Whether all four feet of a horse were off the ground at the same time while trotting?

Difficult for human eye to break down action at fast speed

Videos

- Origin of motion picture takes us to popularly debated question of those times:

Whether all four feet of a horse were off the ground at the same time while trotting?

Difficult for human eye to break down action at fast speed
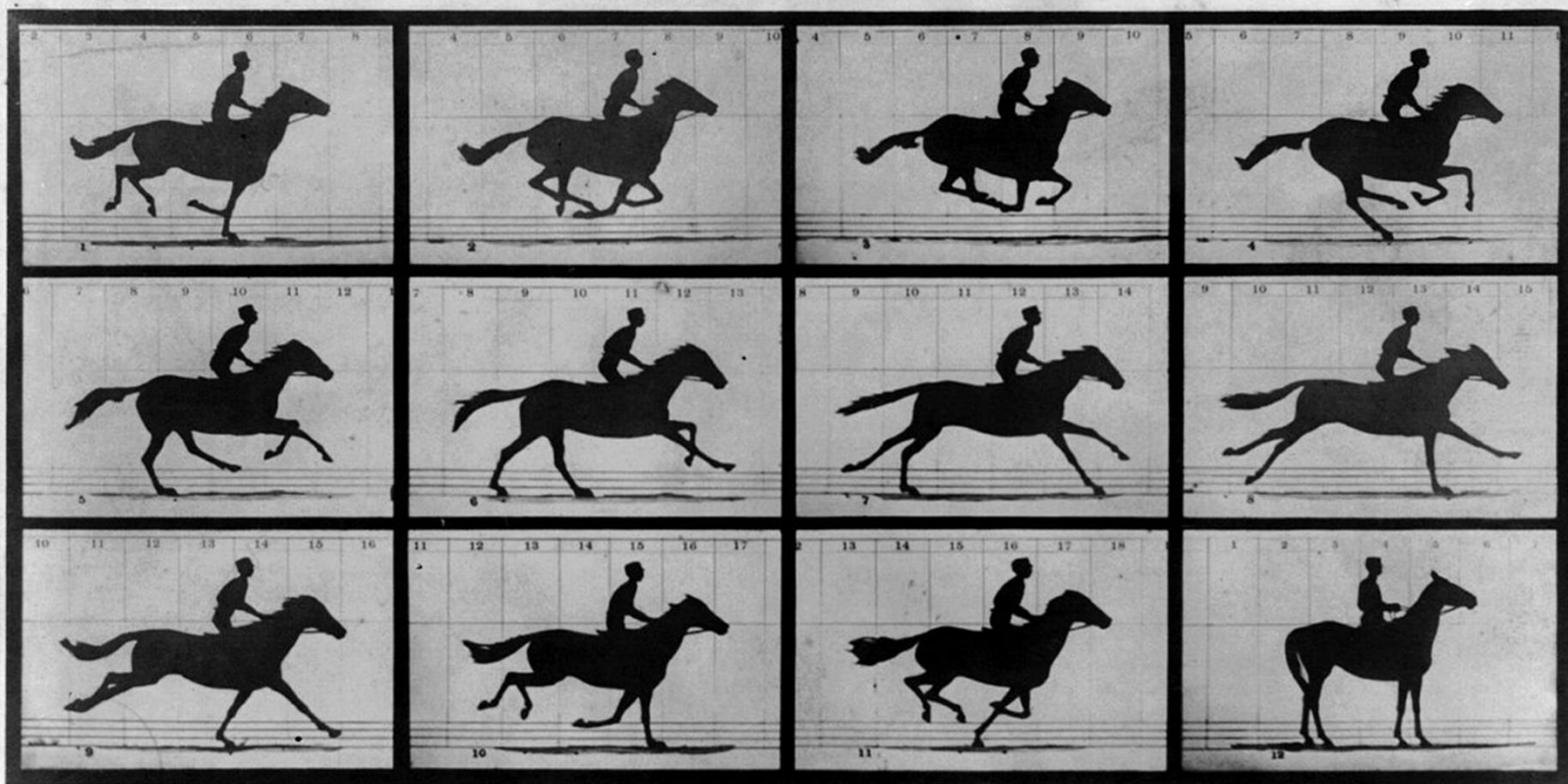
MORSE'S Gallery, 417 Montgomery St., San Francisco.

THE HORSE IN MOTION.

Illustrated by
MUYBRIDGE.

AUTOMATIC ELECTRO-PHOTOGRAPH.

"SALLIE GARDNER," owned by LELAND STANFORD; running at a 1.40 gait over the Palo Alto track, 19th June, 1878.

The negatives of these photographs were made at intervals of twenty-seven inches of distance, and about the twenty-fifth part of a second of time; they illustrate consecutive positions assumed in each twenty-seven inches of progress during a single stride of the mare. The vertical lines were twenty-seven inches apart; the horizontal lines represent elevations of four inches each. The exposure of each negative was less than the two-thousandth part of a second.

Videos
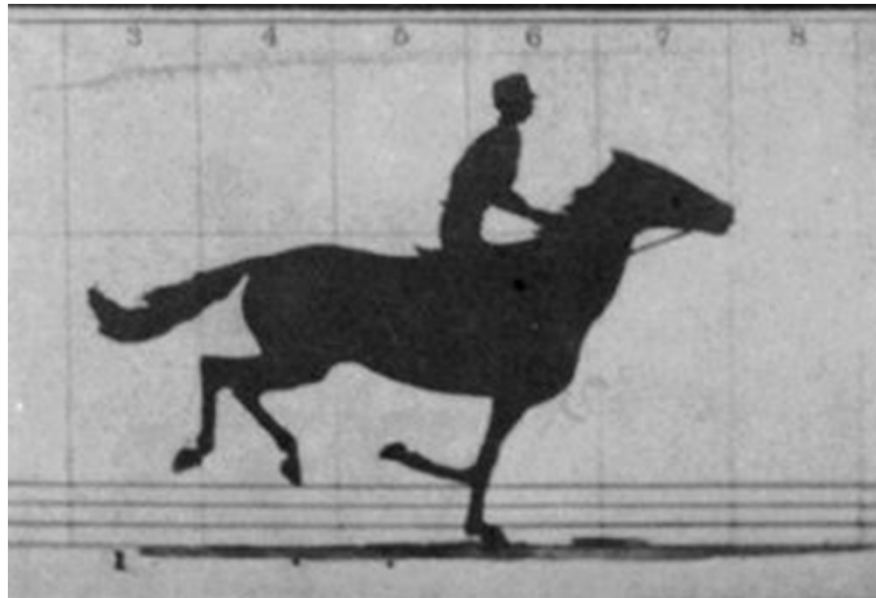
Videos

Videos

Important parameters:

1. Number of frames per second
2. Aspect ratio (for example in TV's previously 4/3, now 16/9)
3. Chroma subsampling (bits per pixel)
4. Compression format (raw, mp4, mpeg etc.)
5. Interlaced vs progressive
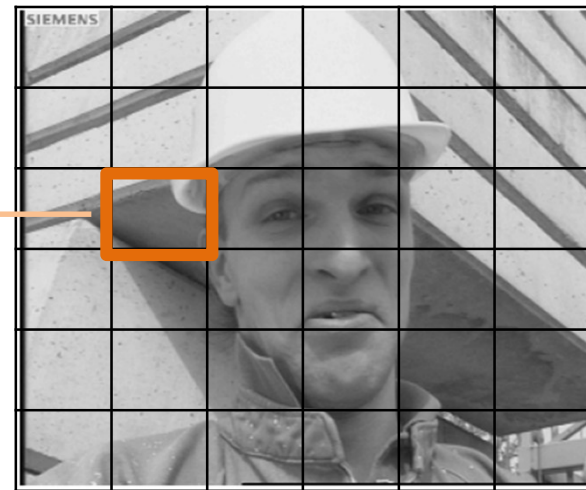
Today's class

- Motion compensation (block matching)
- Video compression

Motion compensation (Block matching)

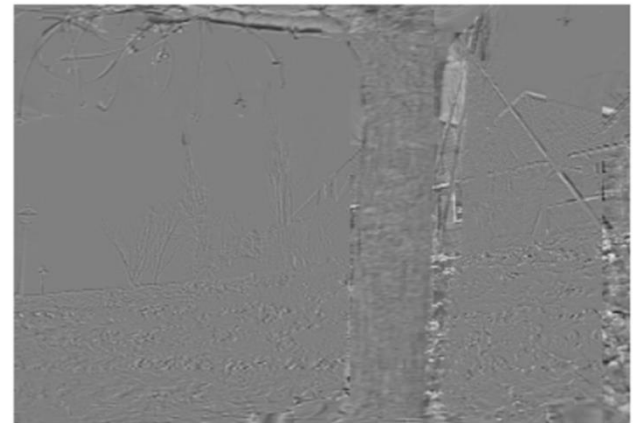frame t-1   frame t

Block matching

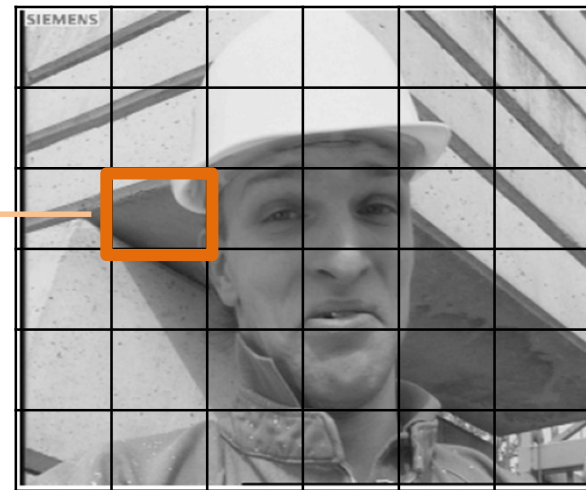# Block matching

Previous frame

Current frame

Current frame with displacement vectors

Motion-compensated Prediction error

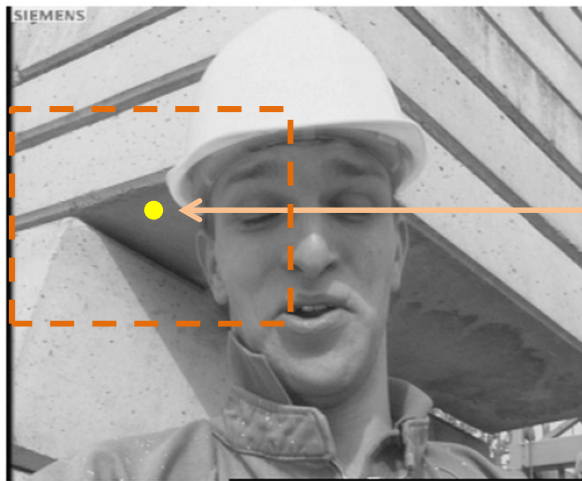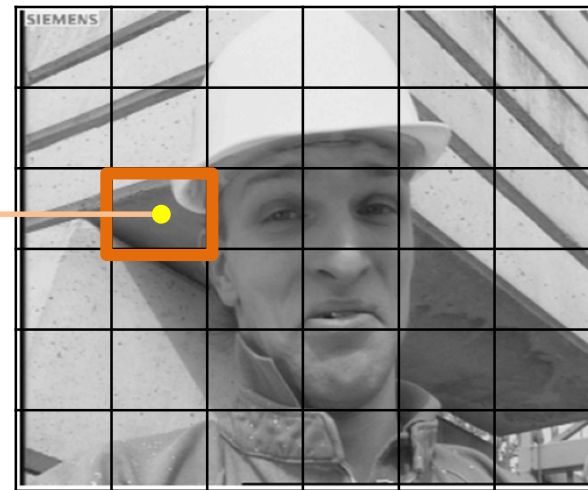# Block matching: How to do it?

frame t-1   frame t

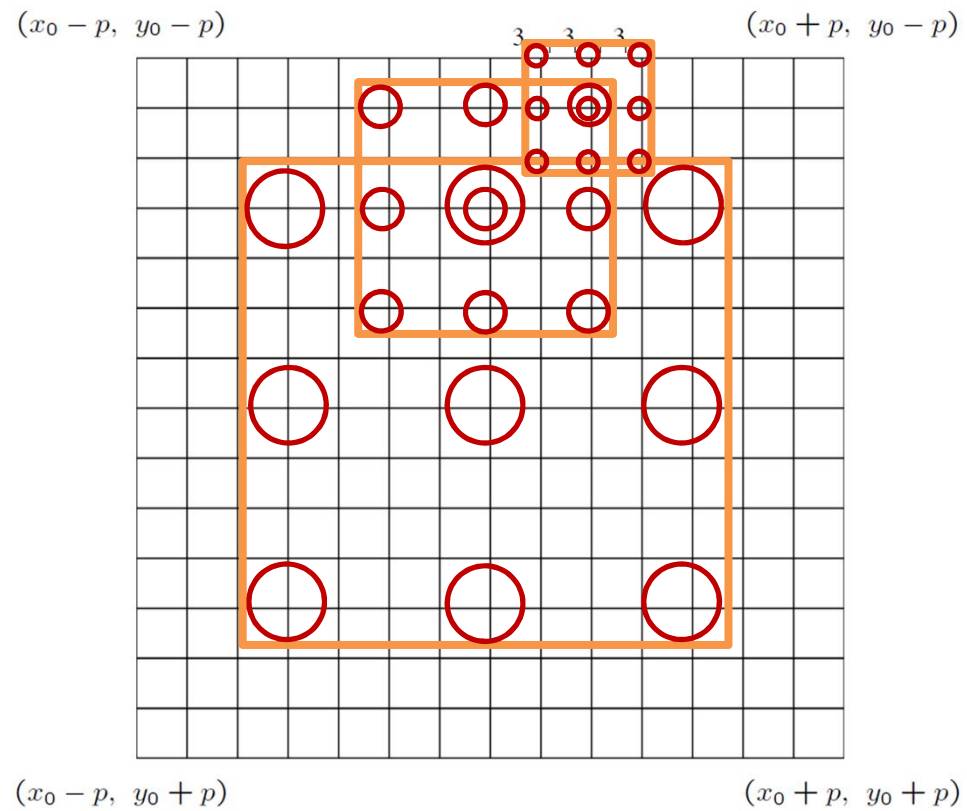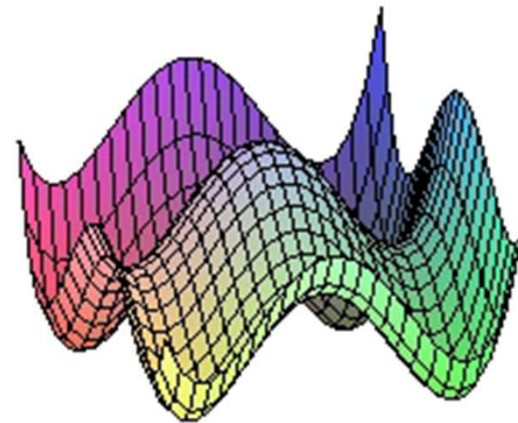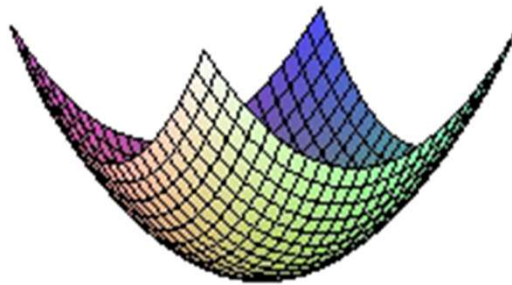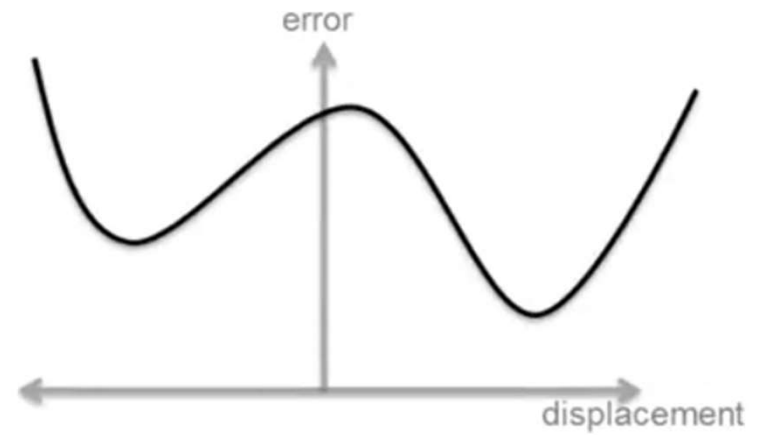# Exhaustive search

frame t-1   frame t

Logarithmic search

Logarithmic search
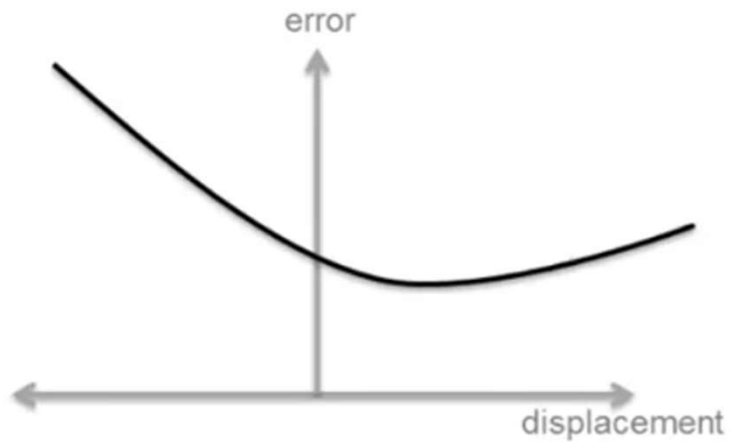
Pixel sub sampling

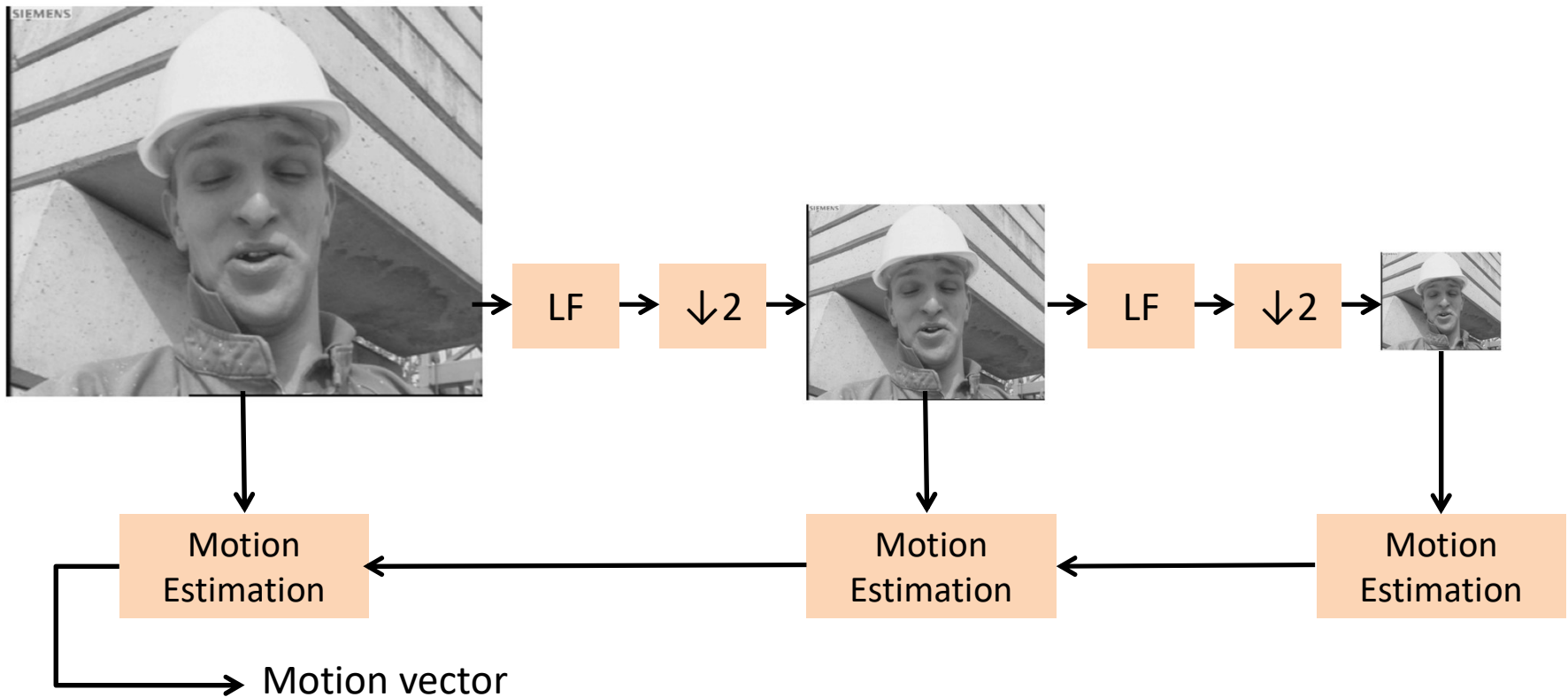# Pixel projection
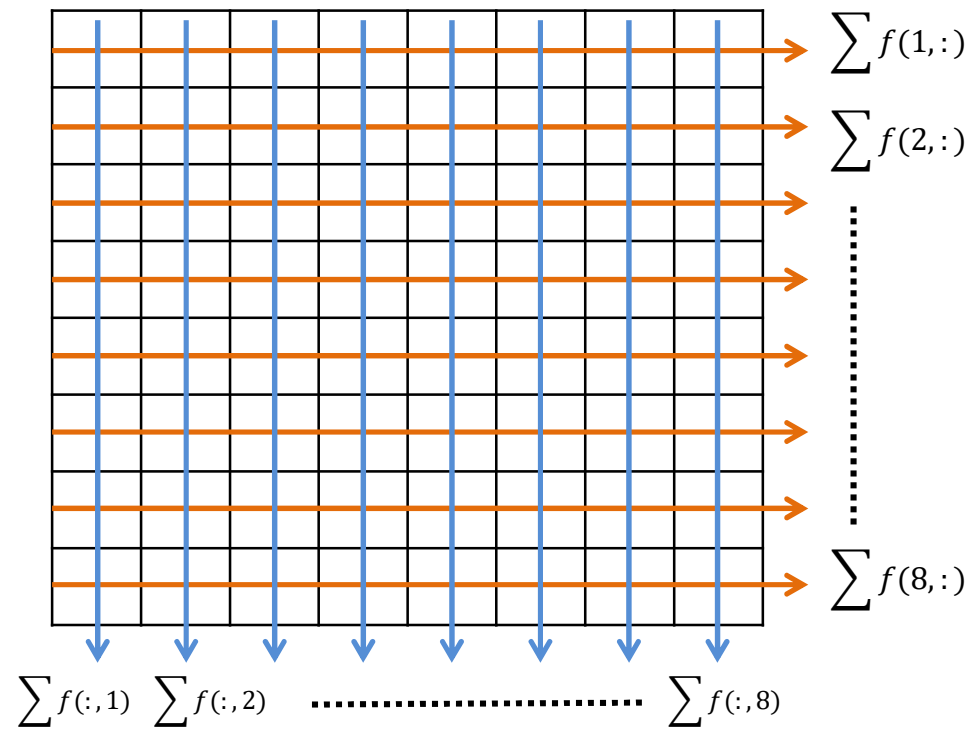
Hierarchical motion estimation

Sub pixel motion estimation

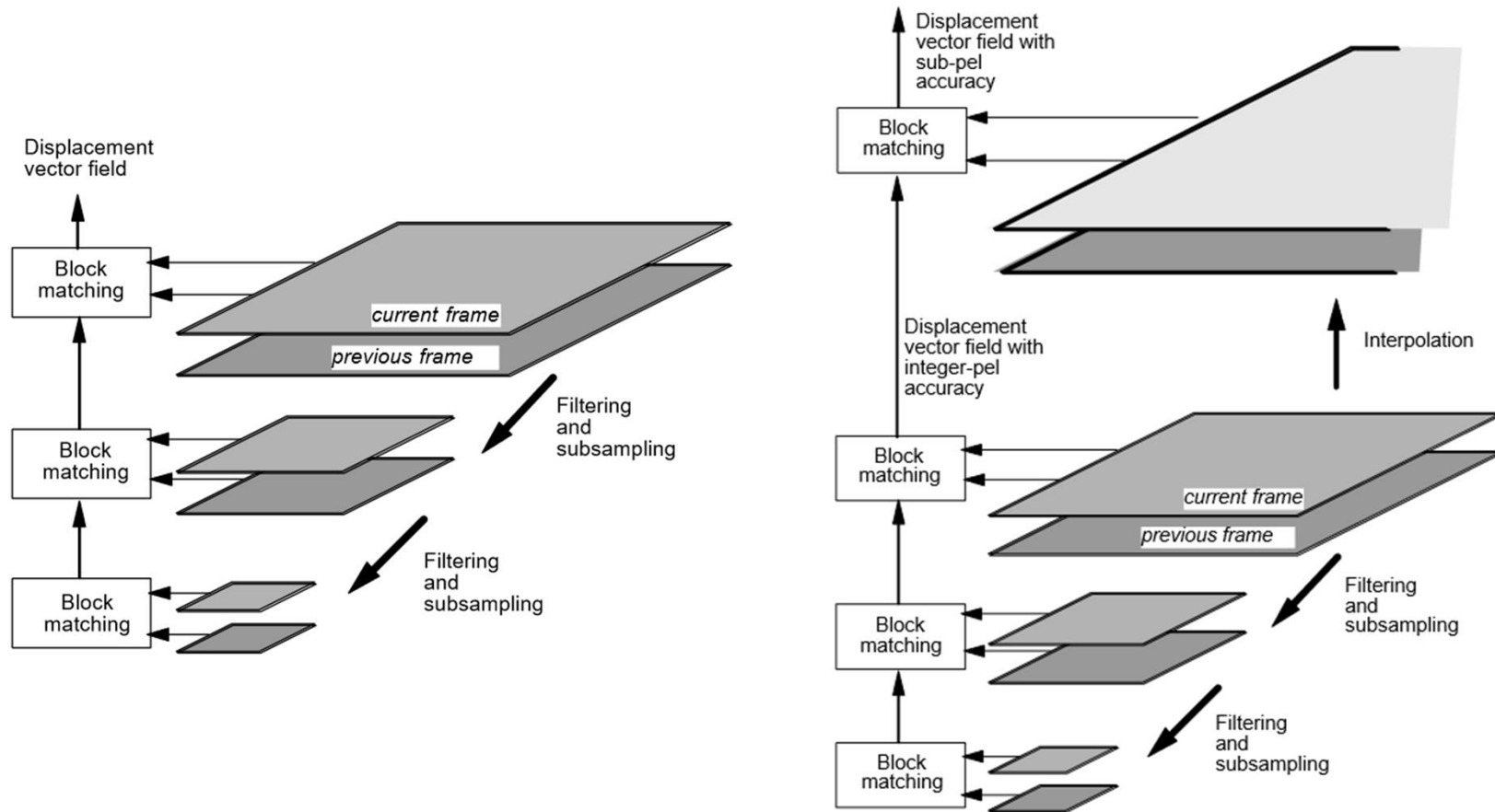Pixel block on half pixel resolution grid

Quarter resolution pixels

Half resolution pixels

Typical Quarter pixel Motion Estimation
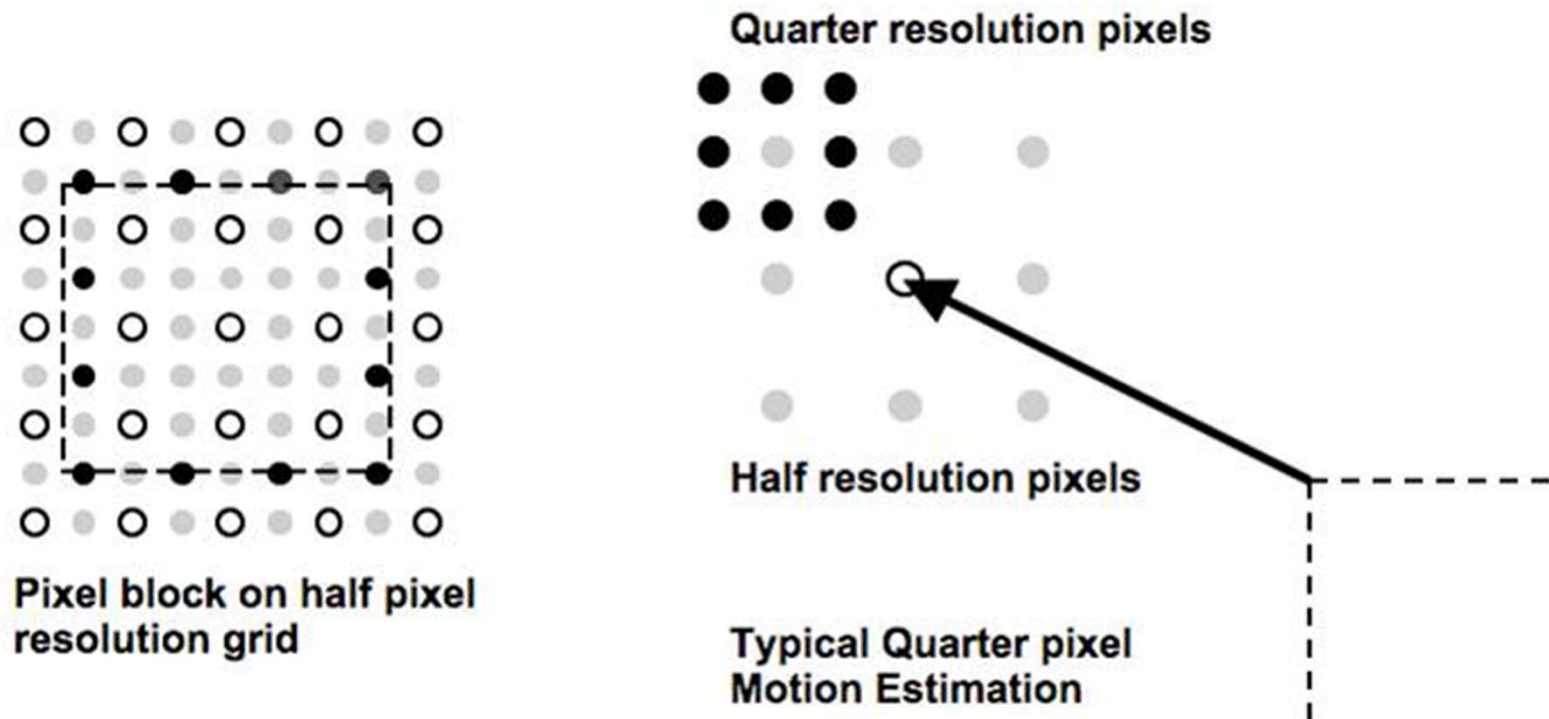
Sub pixel motion estimation (hierarchical view)
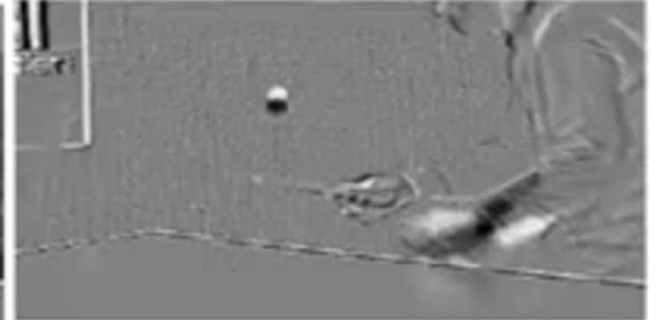
Example results

Reference image   Current image   Frame difference
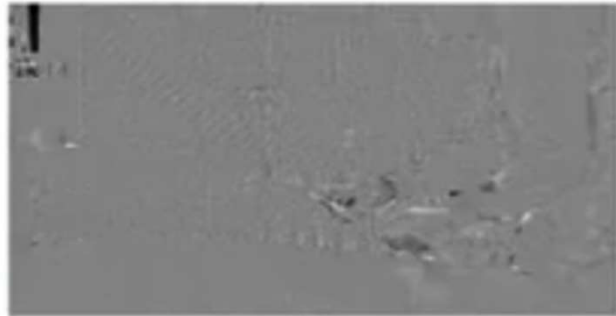
(a)   (b)   (c)
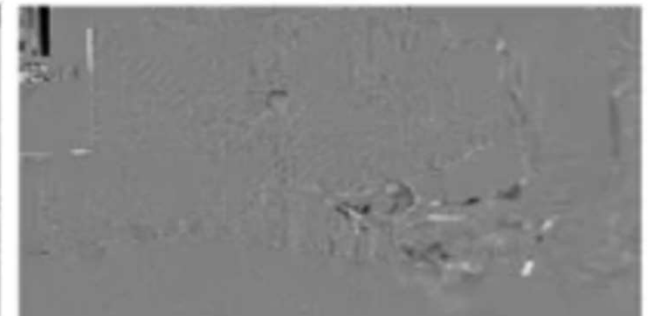
Full search   Logarithmic search   Hierarchical search (3 levels)

(d)   (e)   (f)

VCDemo

Video Compression

- Straight forward solution: take each frame and encode as a jpeg (M-JPEG)
- Can we do better?

Video Compression

Source → **Partition** → **Predict (subtract)** → **Transform** → **Entropy coding** → Compressed video

Quantization (→ Entropy coding)

Compressed video → **Entropy decoding** → **Inverse Transform** → **Predict (add)** → **Reconstruct** → Output
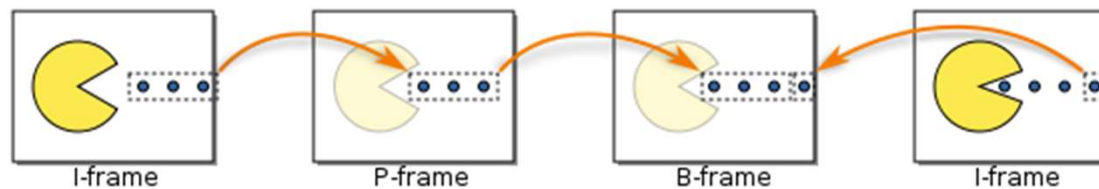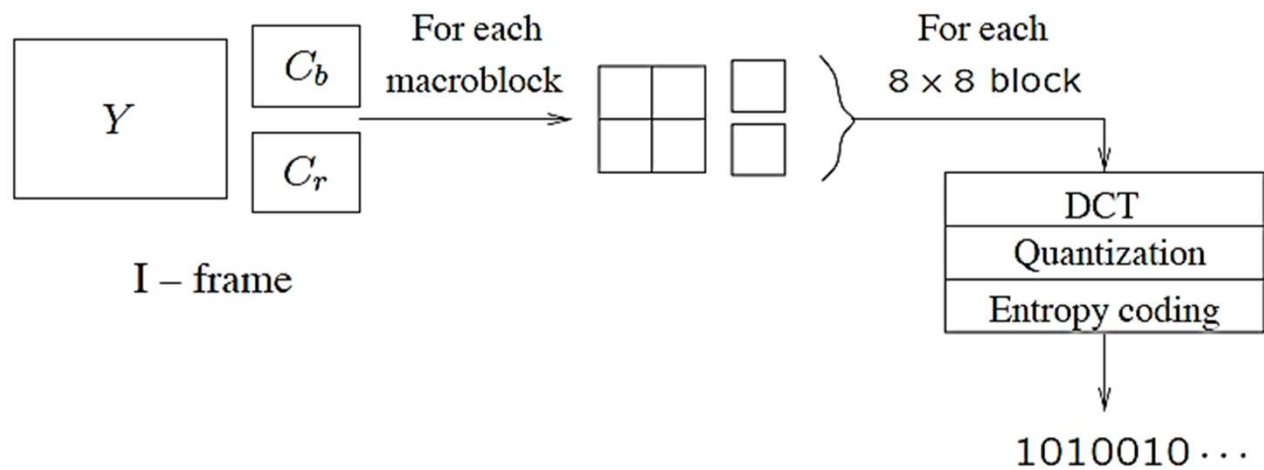
De-quantization (→ Entropy decoding)
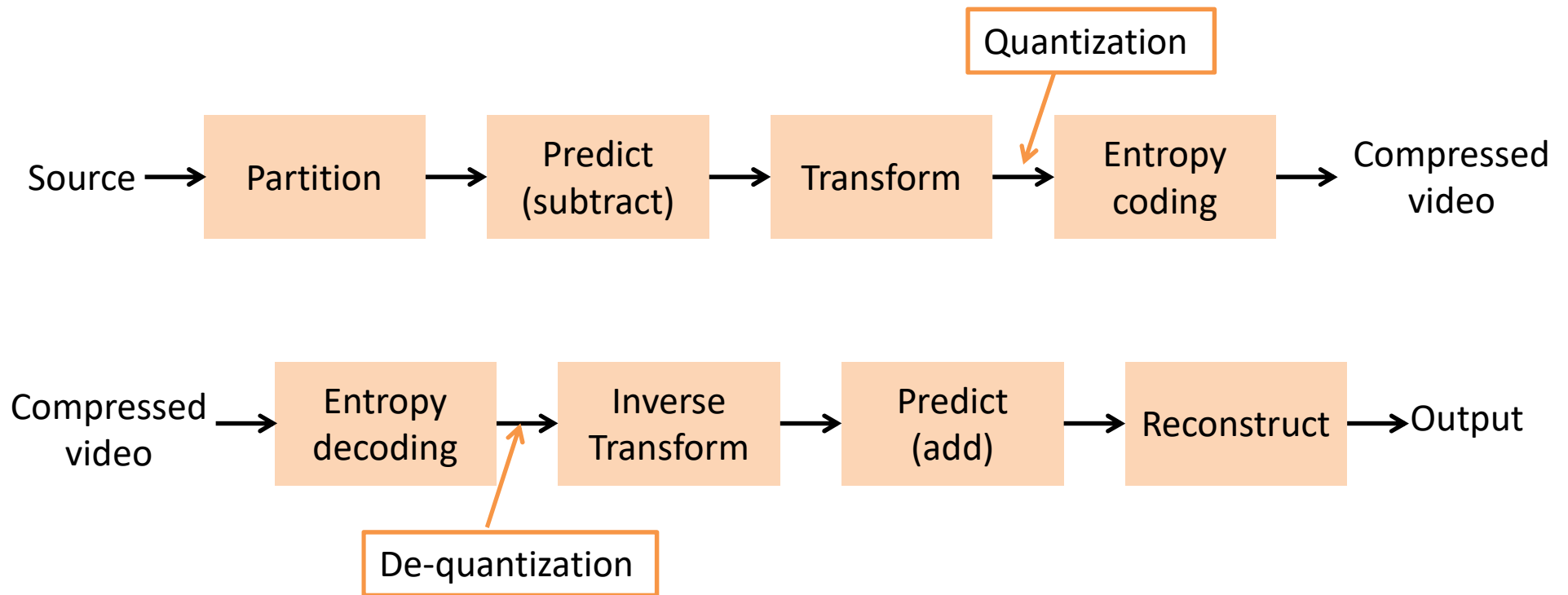
Video compression (type of encoded frames)

- Three frames
  - I frame (intra frame coding)
  - P frame (predicted frame)
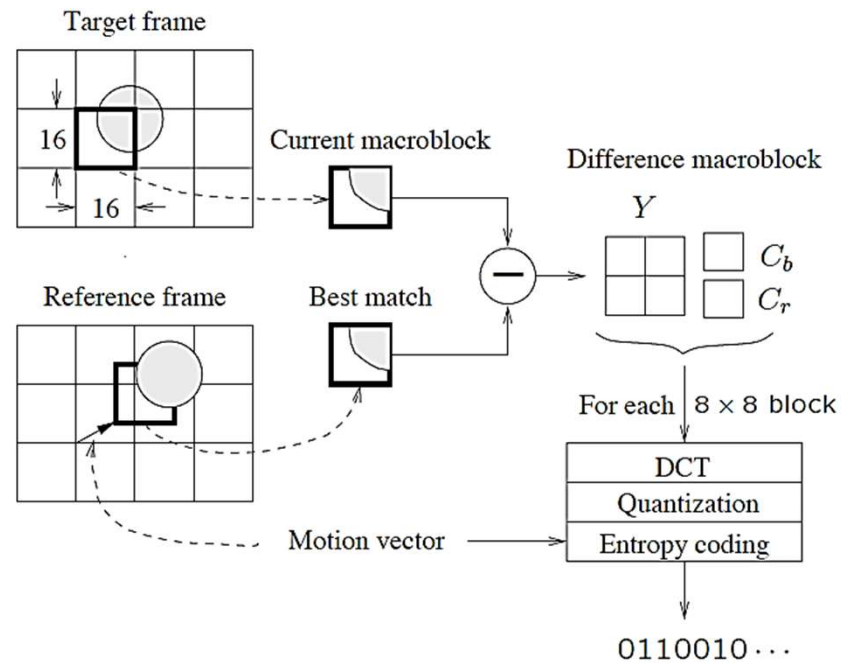 - B frame (bidirectionally interpolated fame)

I-frame   P-frame   B-frame   I-frame

Video Compression

- Intra-frame (I-frame) Coding

Video Compression

- Inter-frame (P-frame) Predictive Coding

Partition and motion estimation

- Assume the current picture can be locally modeled as a translation of the pictures of some previous time.

- Each picture is divided into blocks of 16 x 16 pixels, called a macroblock.

- Each macroblock is predicted from the previous or future frame, by estimating the amount of the motion in the macroblock during the frame time interval

Prediction by motion estimation

Reference frame

Current frame

Residual

Transform + Quantize

Block of samples   After transform   After quantization
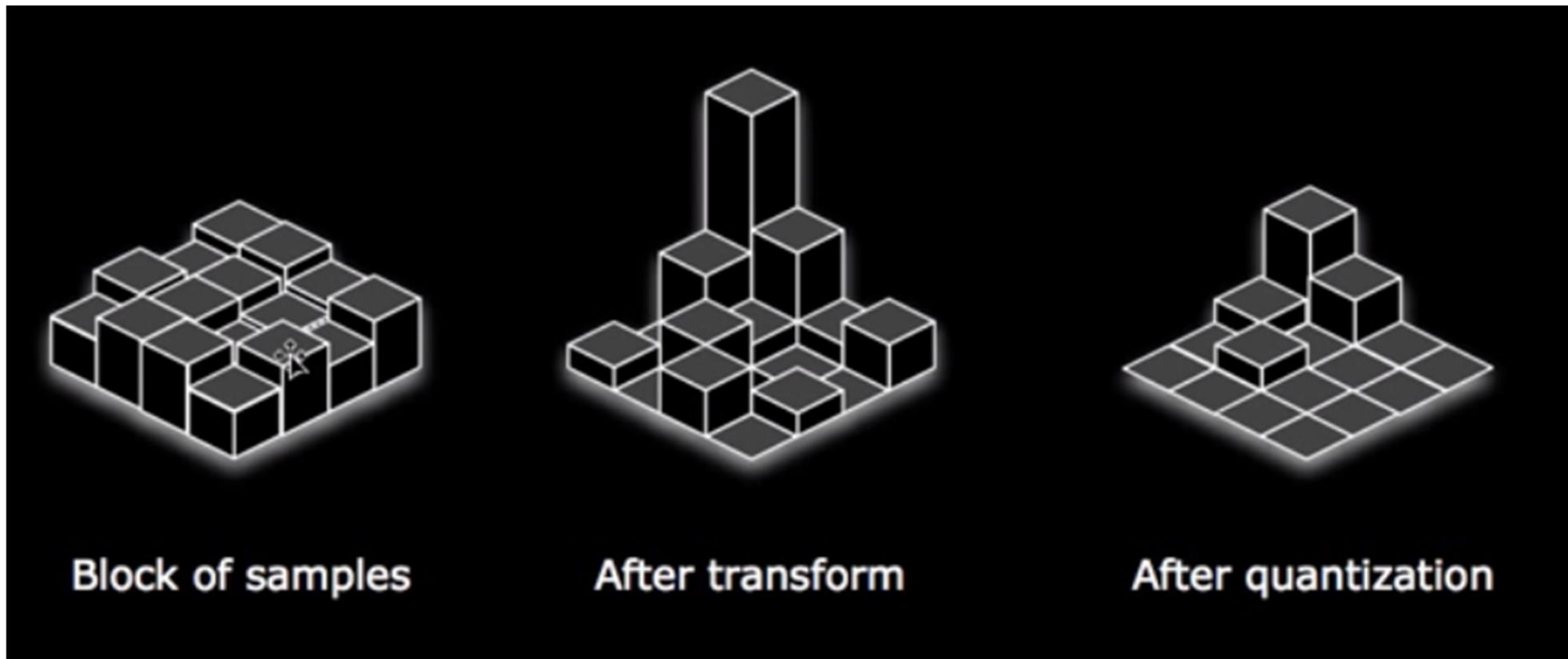
Image source: http://vcodex.blogspot.in/

Entropy coding

- Huffman coding, run length coding etc.

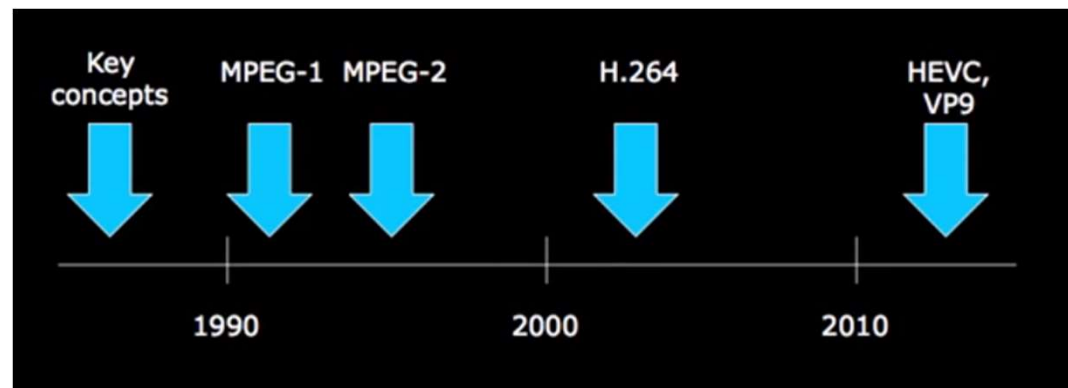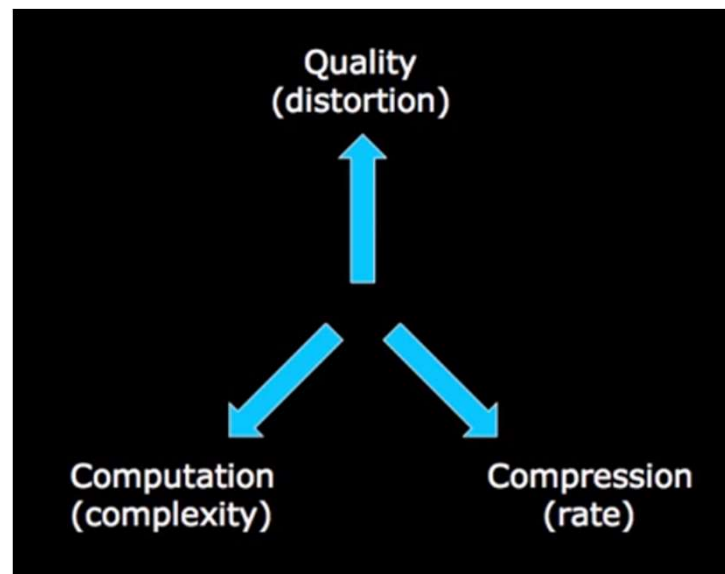# Video compression (type of encoded frames)

MPEG-2

↓ ~2x better

H.264/AVC

Video compression (timeline)

Video compression (trade off)

Salient features H.264

- Variable block size : which block size is better?
  - In terms of number of bits: large is better (less motion vectors need to be computed and encoded)
  - Where is the difficulty?
  - H264 used 16×16 to 4×4 (in fact non square partitions are also allowed)

- Quarter pixel accuracy in motion estimation

- Motion vector over frame boundaries

- Multiple reference frames for prediction (up to 5 previous frames)

- Integer transform (instead of real valued DCT)