Digital Image Processing (CSE 478)
Lecture 21: Motion estimation and video compression

Vineet Gandhi

Center for Visual Information Technology (CVIT), IIIT Hyderabad

Videos

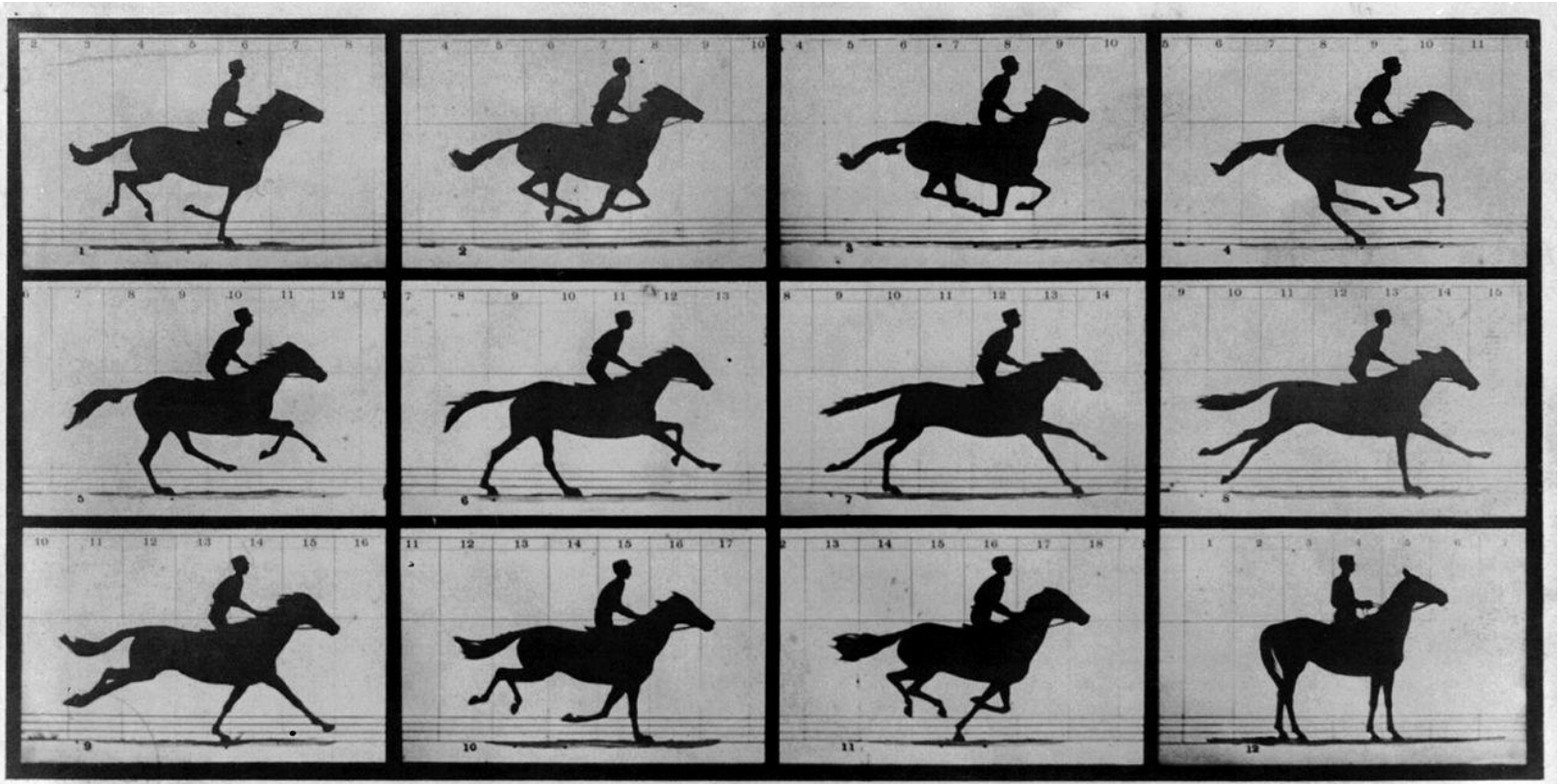
- A sequence of still frames shown together

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



MORSE'S Gallery, 417 Montgomery St., San Francisco.

THE HORSE IN MOTION.

Illustrated by

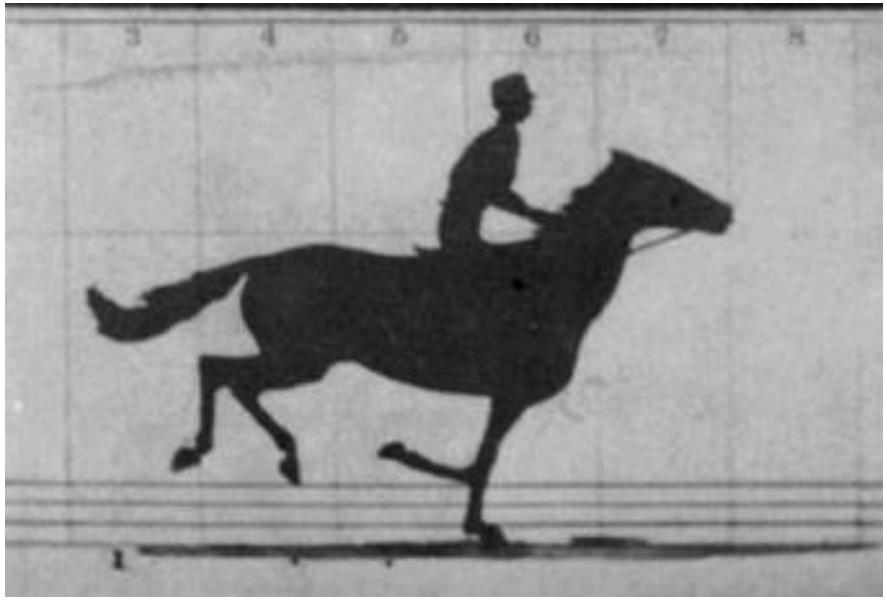
MUYBRIDGE.

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

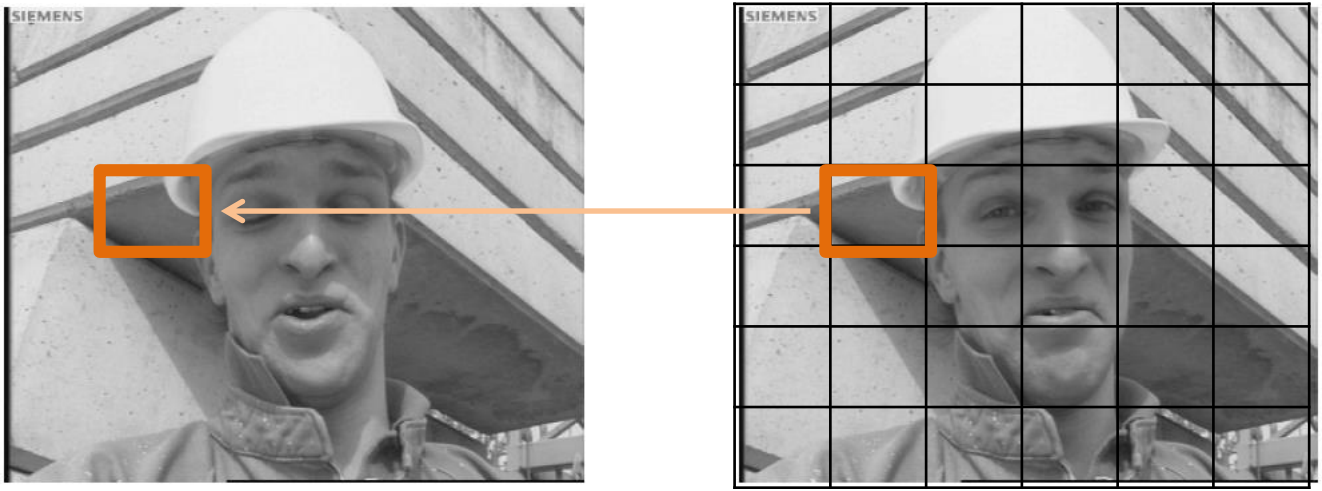
Important parameters:

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)

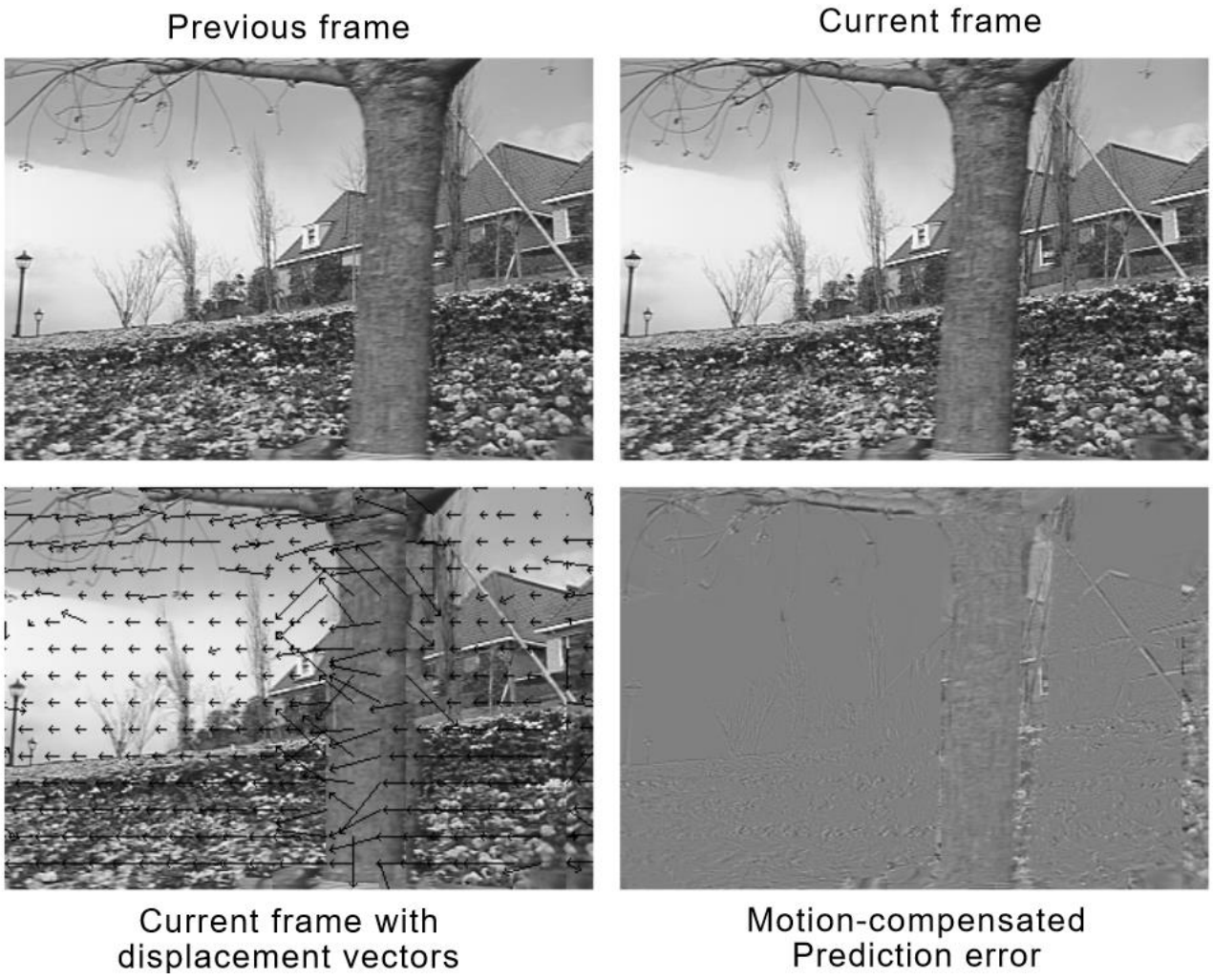


frame t-1

frame t

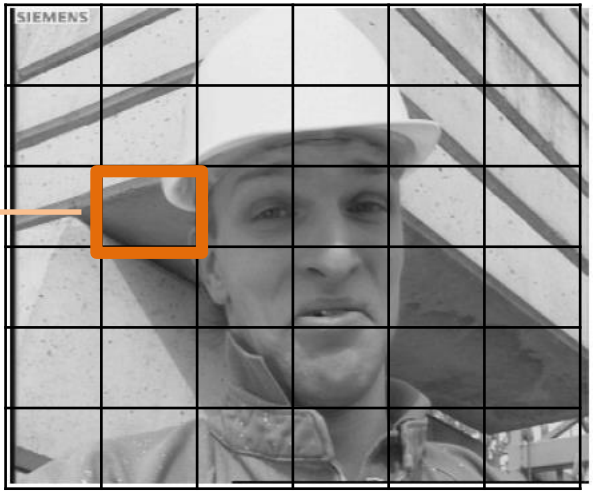
Block matching

Block matching



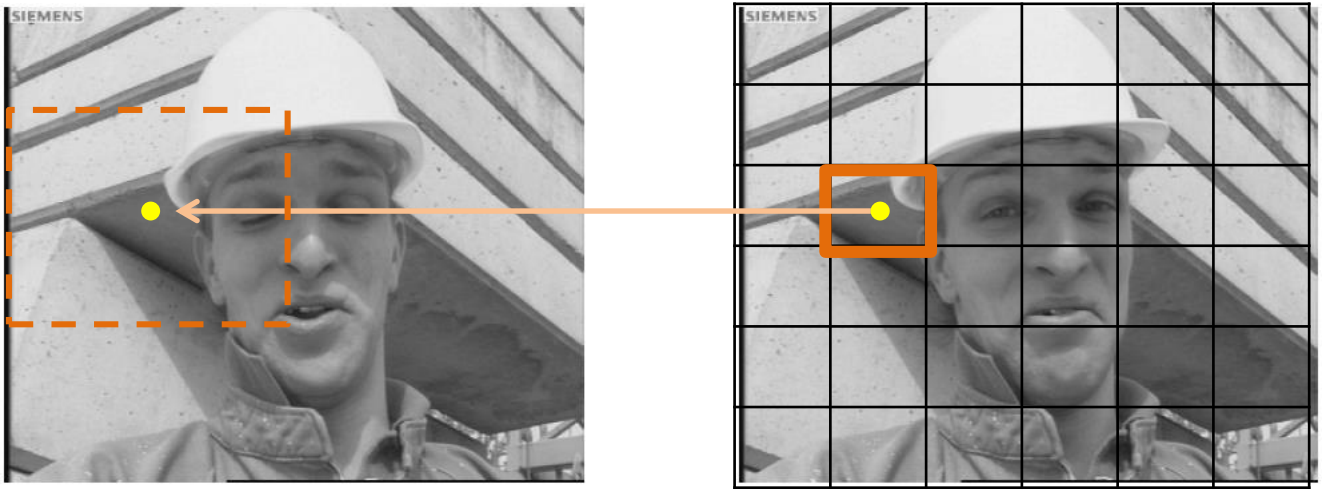
Block matching: How to do it?

frame t-1



frame t

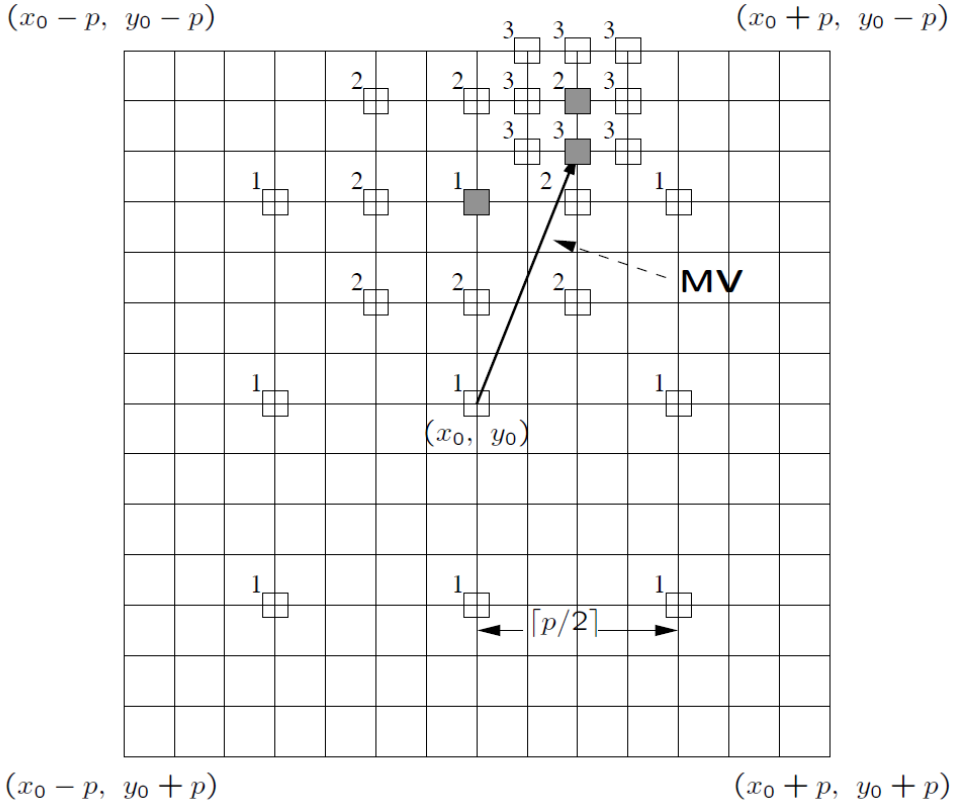
Exhaustive search



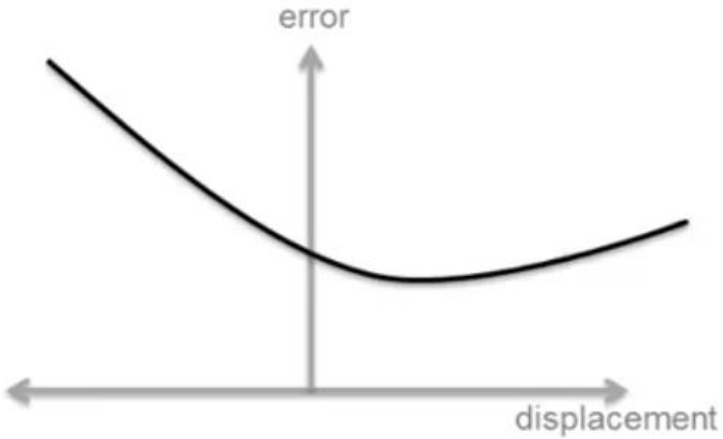
frame t-1

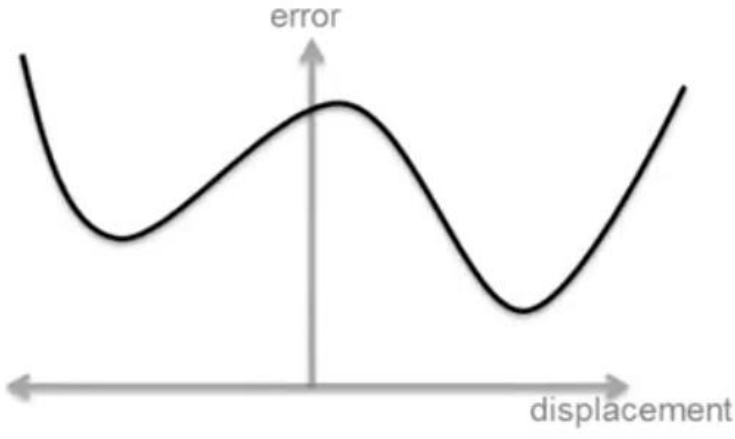
frame t

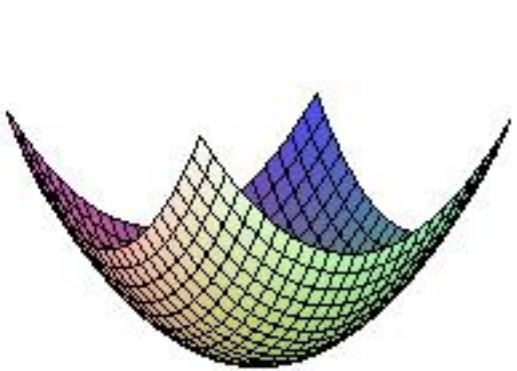
Logarithmic search

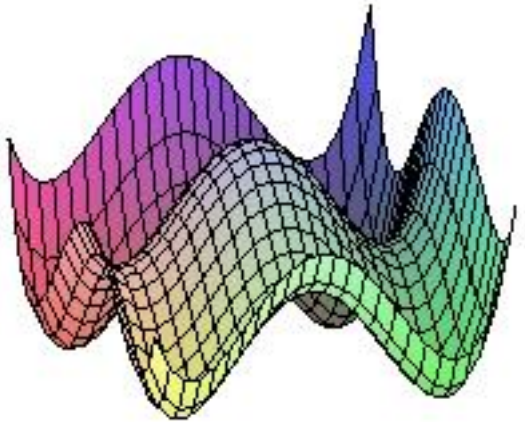


Logarithmic search

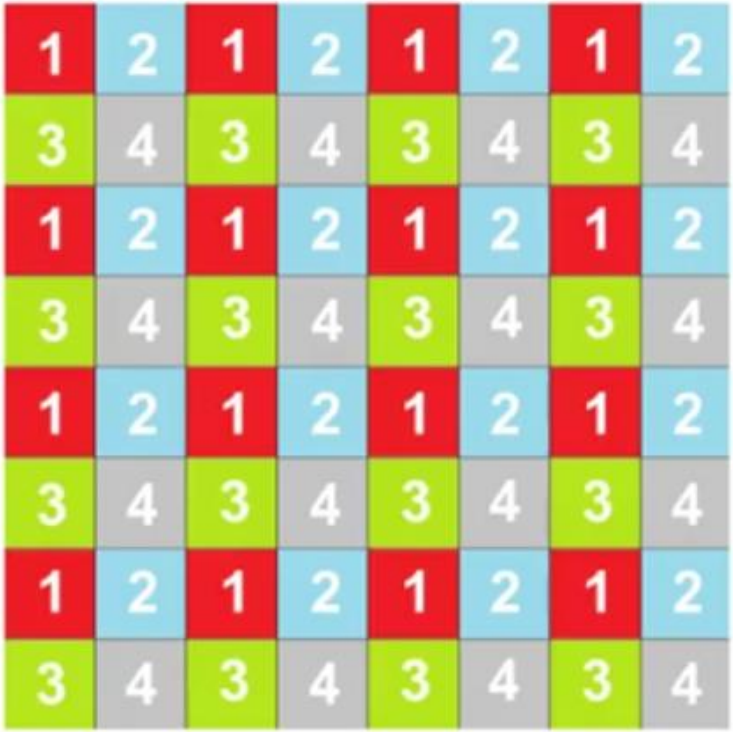




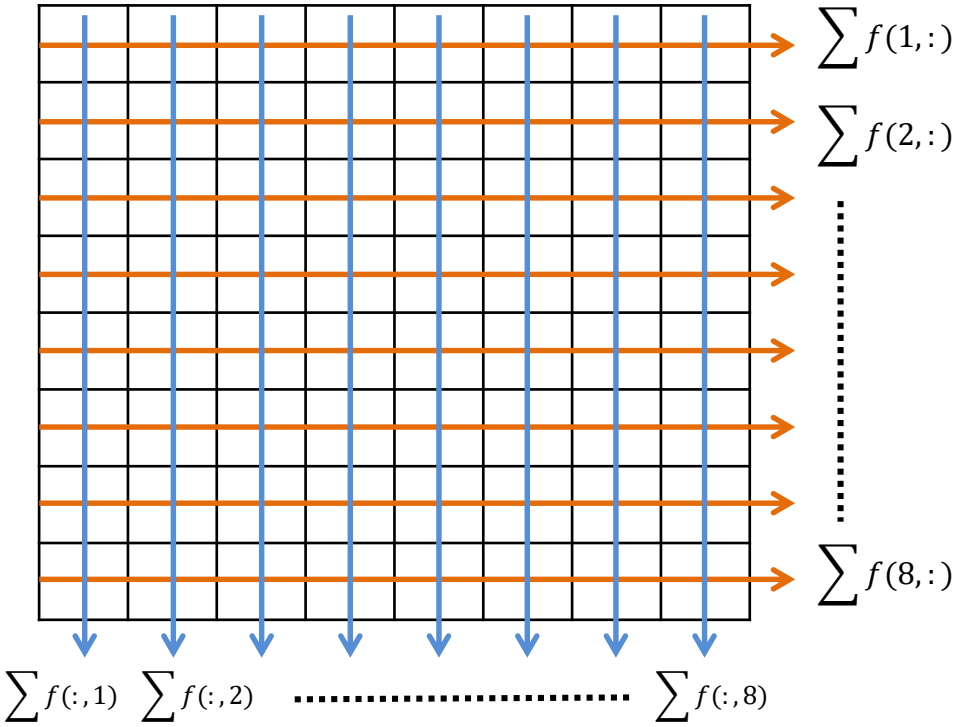




Pixel sub sampling



Pixel projection



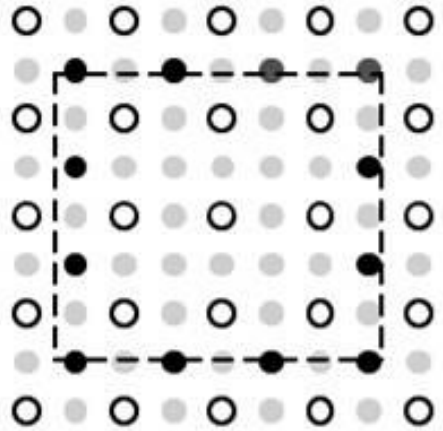
Hierarchical motion estimation

LF → ↓2 → LF → ↓2

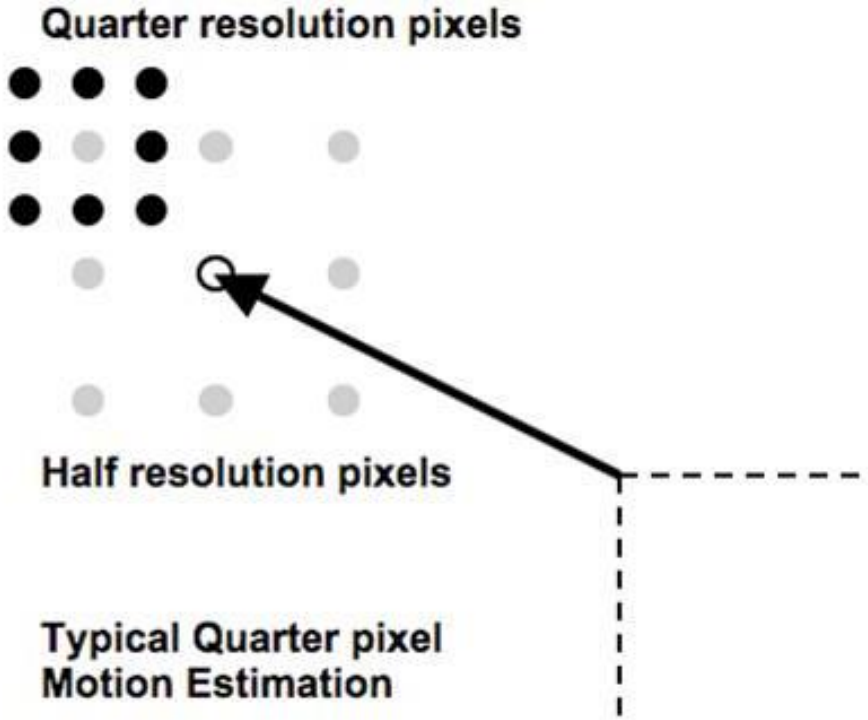
Motion Estimation ← Motion Estimation ← Motion Estimation

Motion vector

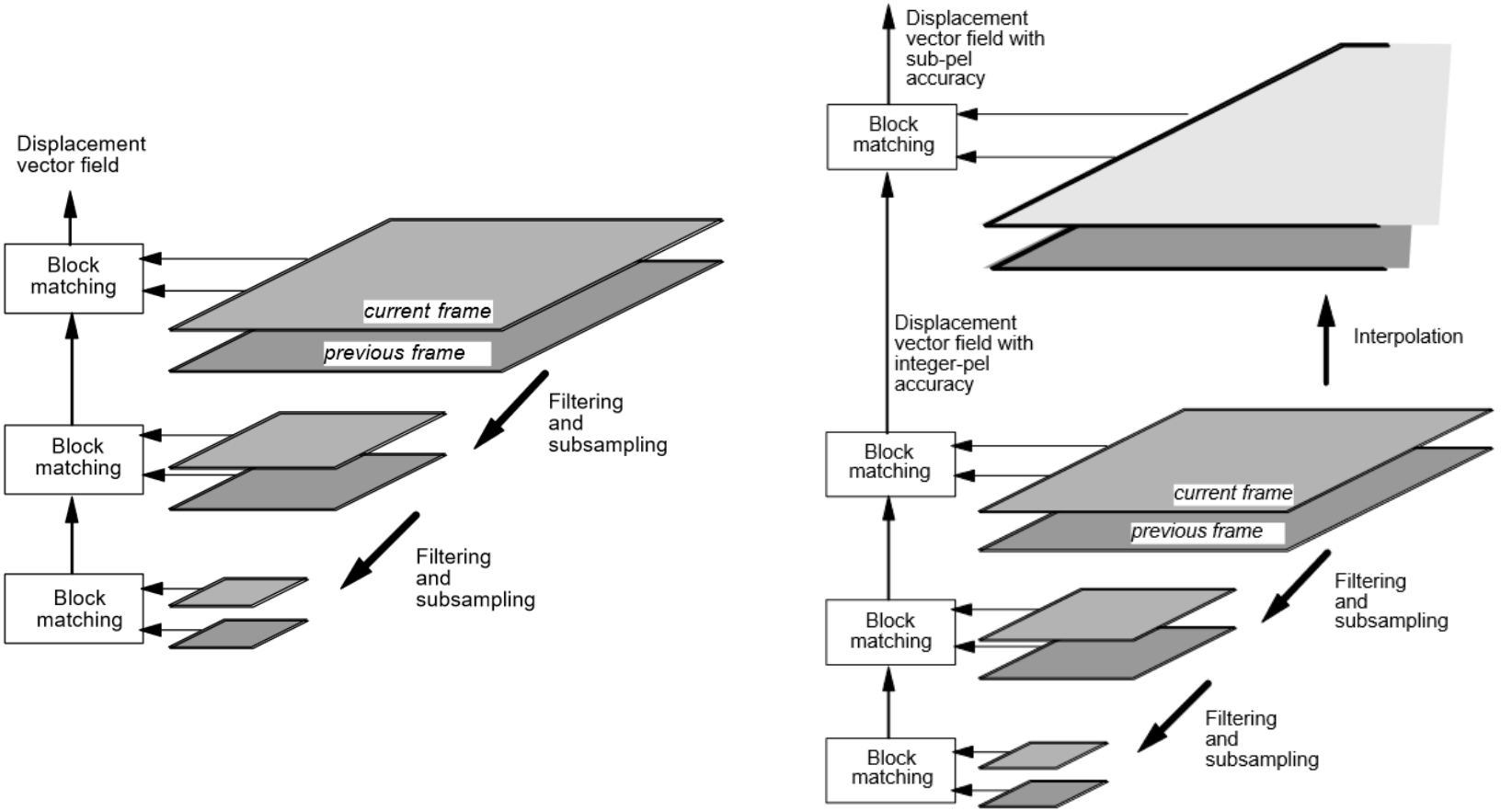
Sub pixel motion estimation



Pixel block on half pixel resolution grid



Sub pixel motion estimation (hierarchical view)



Example results

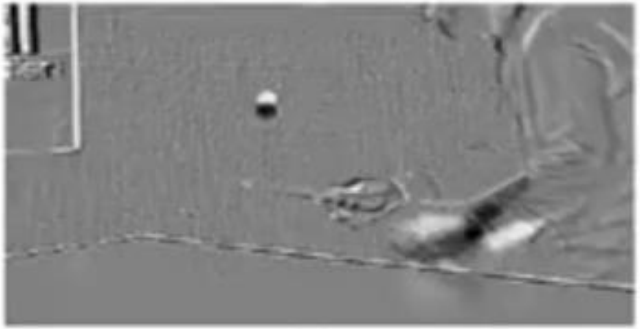
Reference image

(a)

Current image

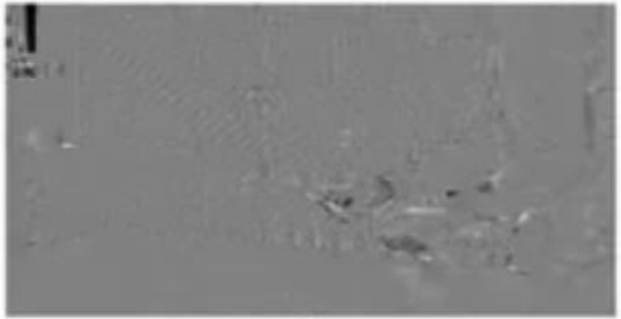
(b)

Frame difference



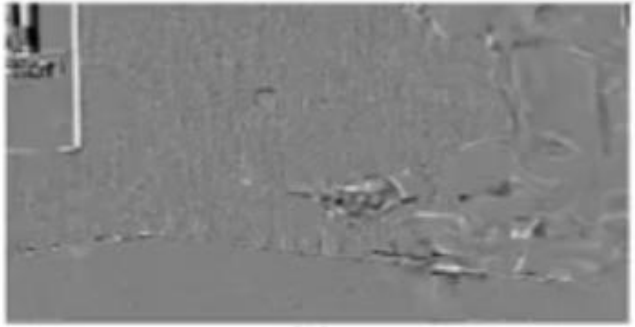
(c)

Full search



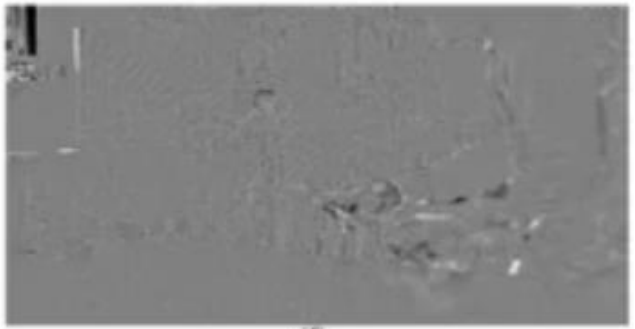
(d)

Logarithmic search



(e)

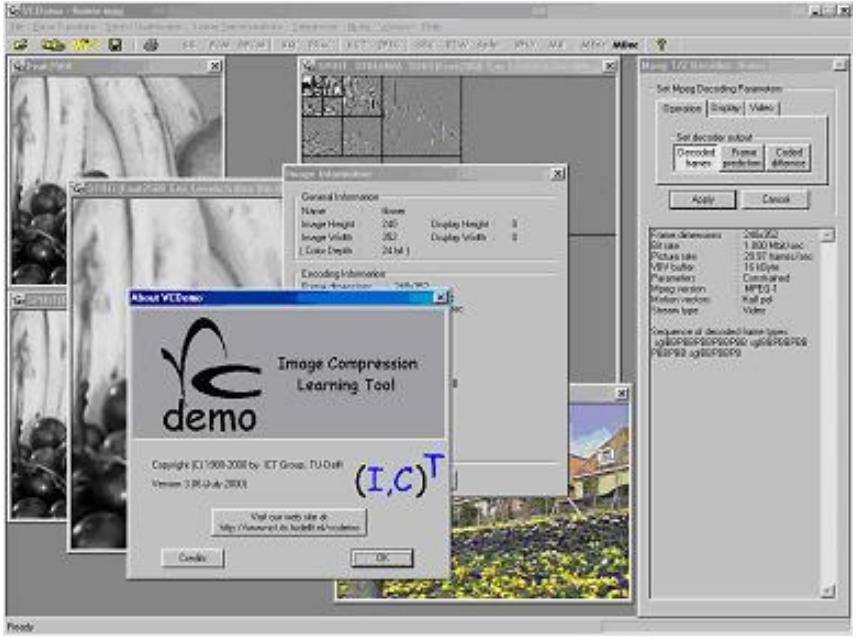
Hierarchical search (3 levels)



(f)

Courtesy: V. Bhaskaran and K. Konstantinides, Image and Video Compression Standards: Algorithms and Architectures , Springer, 1997

VCDemo



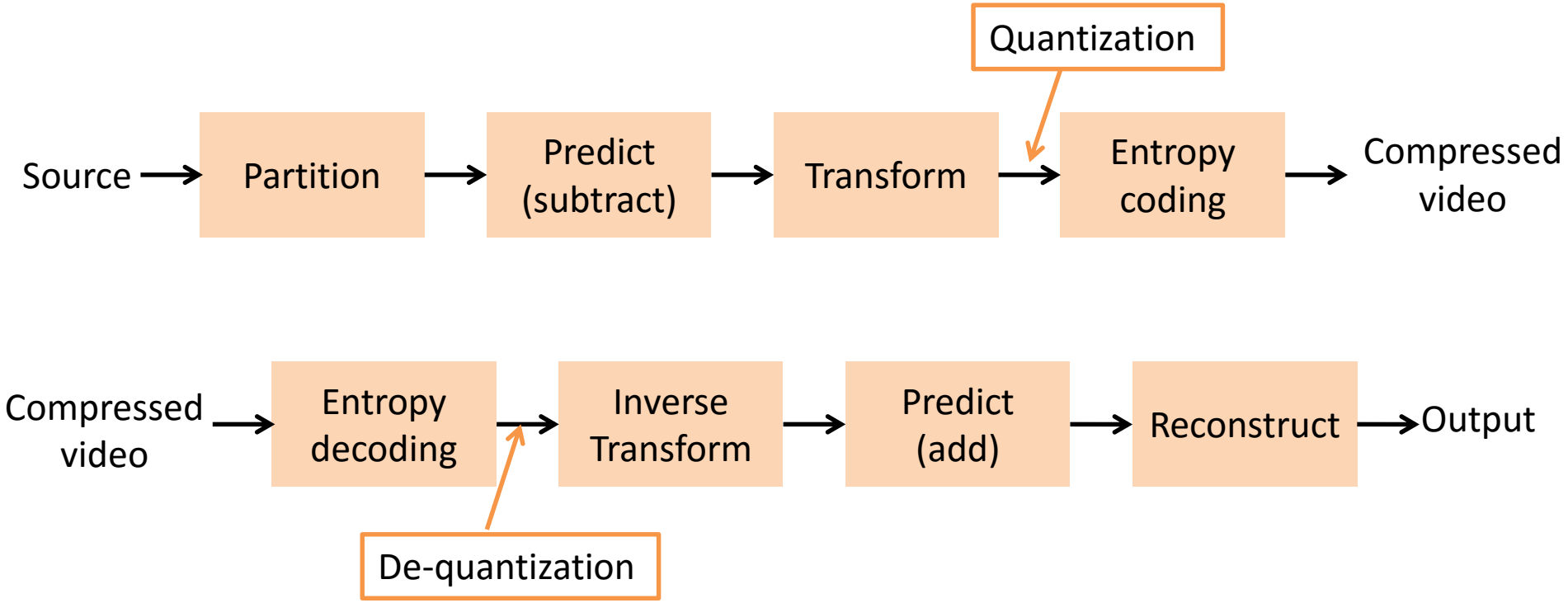
Video Compression

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

Frame 66

Frame 69

Video Compression



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

Image source: http://vcodex.blogspot.in/

Transform + Quantize

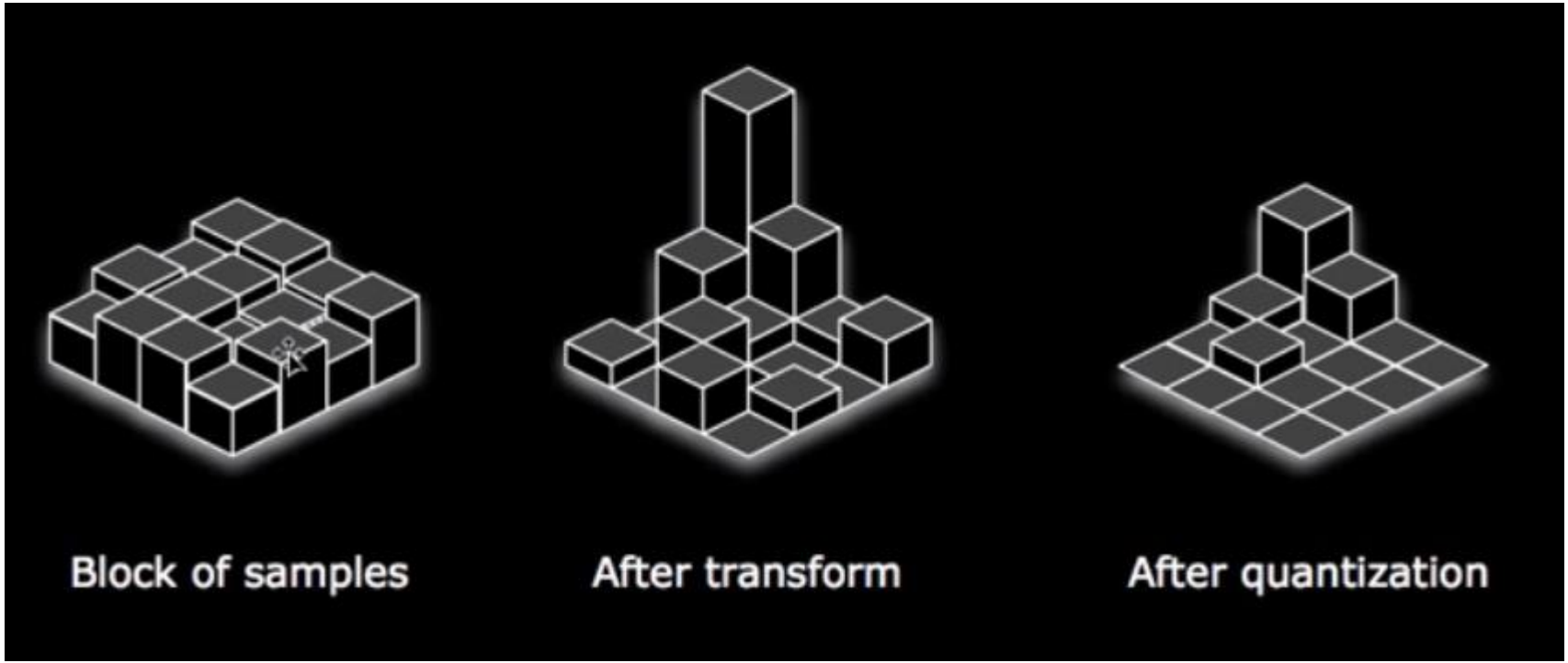


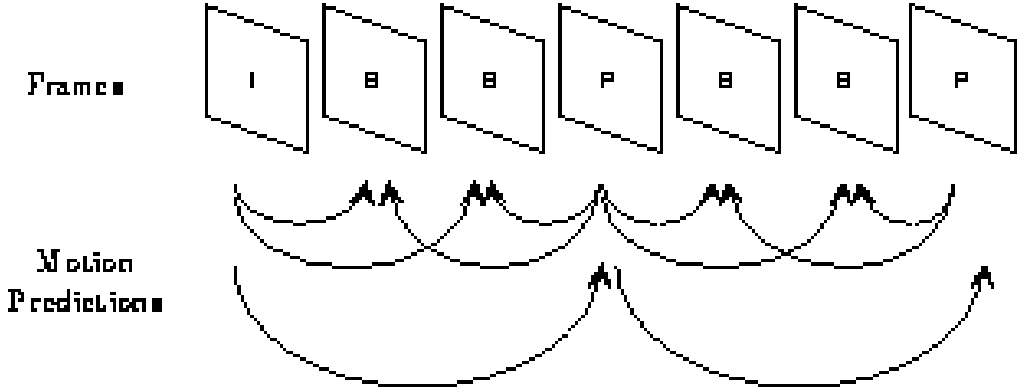
Image source: http://vcodex.blogspot.in/

Entropy coding

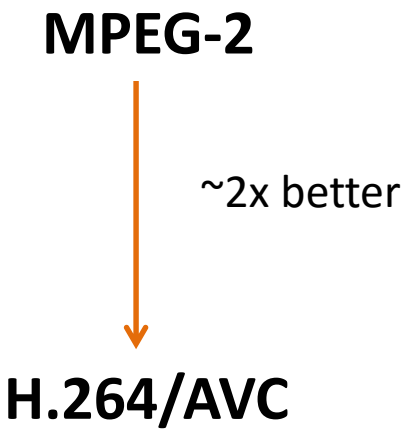
- Huffman coding, run length coding etc.

Video compression (type of encoded frames)

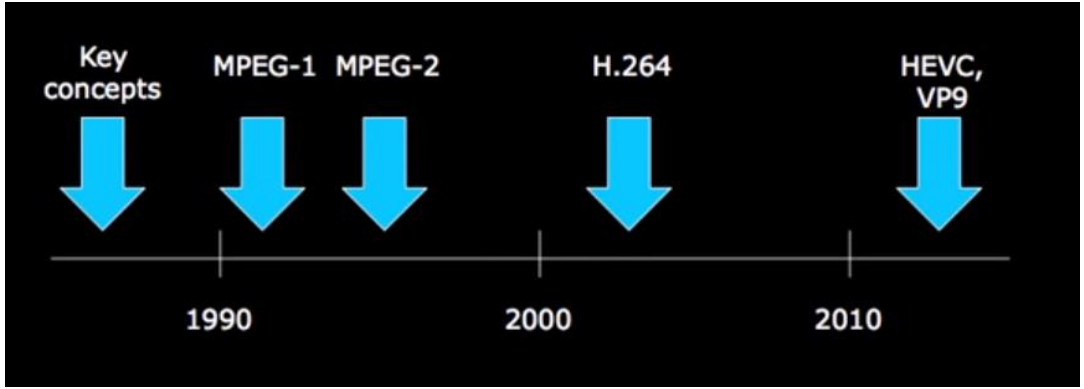
- Three frames
  - I frame (intra picture)
  - P frame (predicted picture)
  - B frame (bidirectionally interpolated picture)



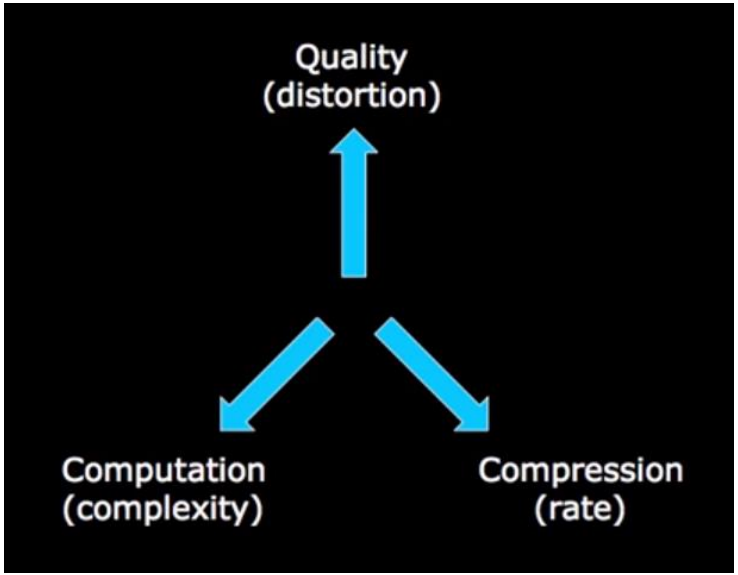
Video compression (type of encoded frames)



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: small 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)