Video Compression Concepts

Nimrod Peleg
<u>Update</u>: March. 2009

The need for compression

- Typical Movie ≈ 150 GBytes
  - 100 minutes * 60 sec/min * 30 frames/sec * 525 rows * 525 columns/row * 24 bits/pixel ≈ 1200 Gbits ≈ $150*10^9$ Bytes
- DVD can hold only 4.7 GBytes
  - Compression is a must!

- Minor noticeable difference
- More than 1:10 for image and up to 1:100 for video

- Compression in usually lossy

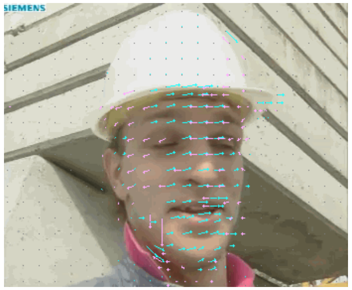
Sample Applications

- Most images and video sequences are stored in compressed form

- Many DSP algorithms are more efficient when implemented in the compressed domain

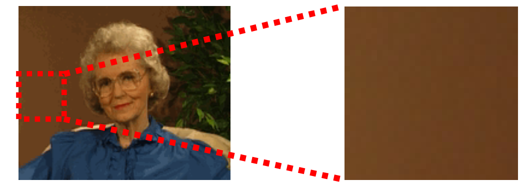
<u>For example</u>

- Face recognition & tracking
- Motion detection
- Image & video retrieval
- Logo insertion
- Transcoding



Redundancy ⇔ Irrelevancy

- Redundancy
  - ✓**Spatial**: pixel-to-pixel or spectral correlation



  - ✓**Temporal**: frame-to-frame similarity

N N+1 N+2

  - ✓**Statistical**: non-uniform distribution of data

Luminance Histogram

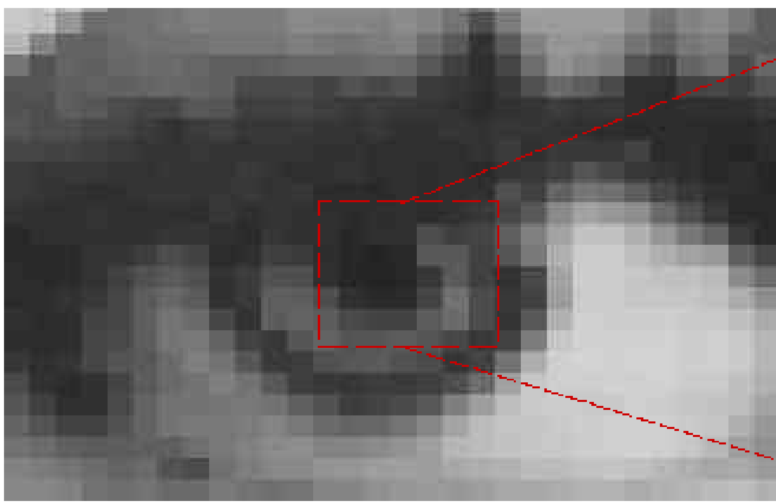
- Irrelevancy relates to an **observer viewing** an image

Redundancy + Irrelevancy ➲ high compression ratio

loss

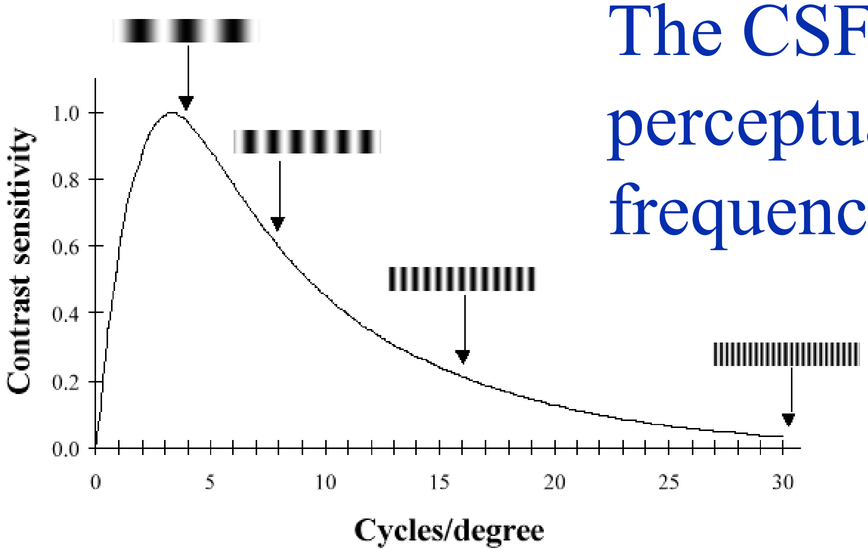
Spatial Redundancy & Irrelevancy

- What is the value of the missing pixel? It is 39.
- How critical is it to correctly reproduce it?

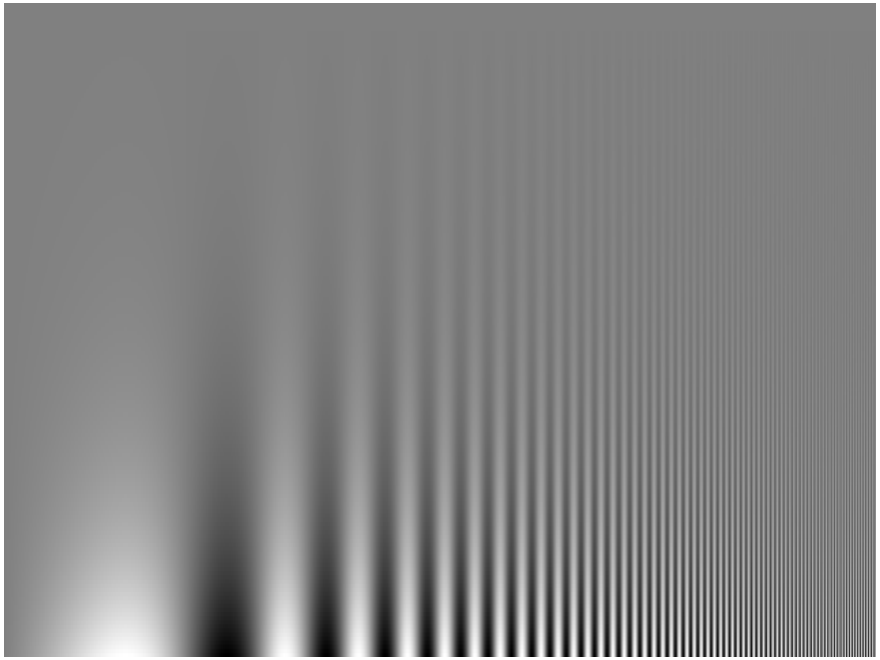


46	44	46	47	44	50	56
50	40	46	48	76	62	69
45	39	36	40	88	87	65
48	38	36	?	56	90	65
76	48	41	43	69	112	77
90	72	66	66	90	108	74
98	86	86	91	83	72	57

Contrast Sensitivity Function



The CSF illustrates the limited perceptual sensitivity to high spatial frequencies



Visual Masking

original

distortion in smooth area

distortion in busy area

Visual Masking

original

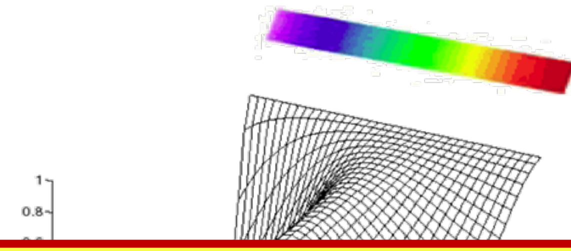
distortion in smooth area

distortion in busy area

Video Compression enablers

Video clips are:

- Spatially redundant



Temporal redundancy

Human eyes are:

- Less sensitive to high spatial frequencies
- Less sensitive to chromatic resolution
- Less sensitive to distortions in “busy” areas

Interframe Compression

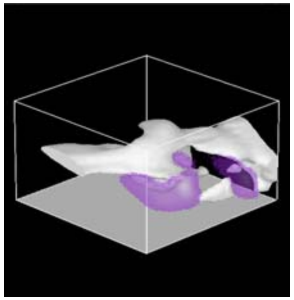
- The simple way is a frame-by-frame compression (JPEG-like):

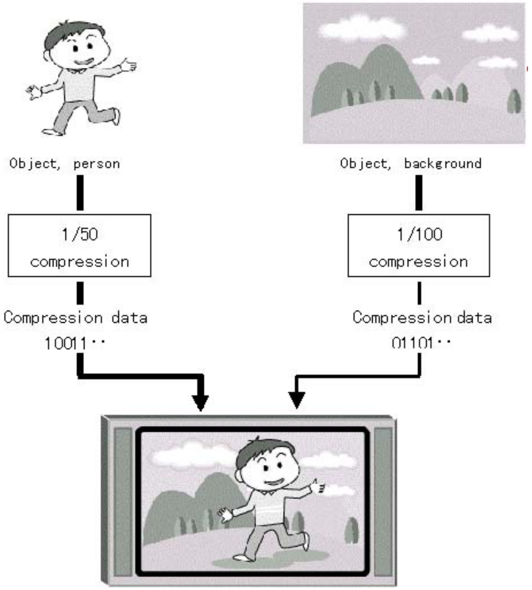
  Limited compression ratio

- Interframe compression exploits temporal redundancies : similarity between successive frames
  - In addition to spatial, spectral and psychovisual redundancies in still image compression)
- No Random Access at all frames !

Possible Strategies

- 3-D Waveform coding
  (TC / Sub-Band)
- Object/Knowledge based coding
- Motion Compensation based coding





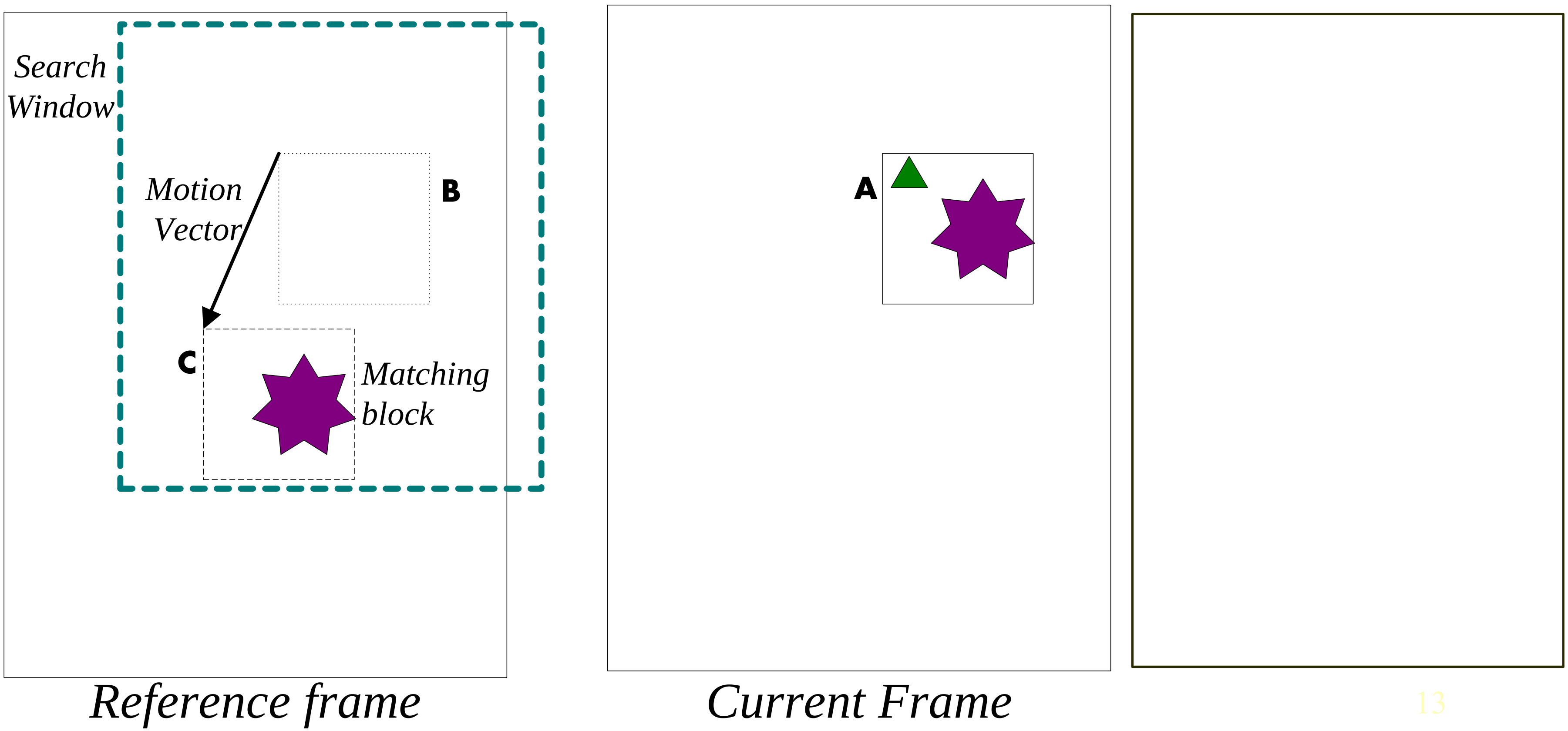
Motion Compensation Approach

- Segmenting each frame to "changed" and "unchanged" regions (Vs. previous frame)

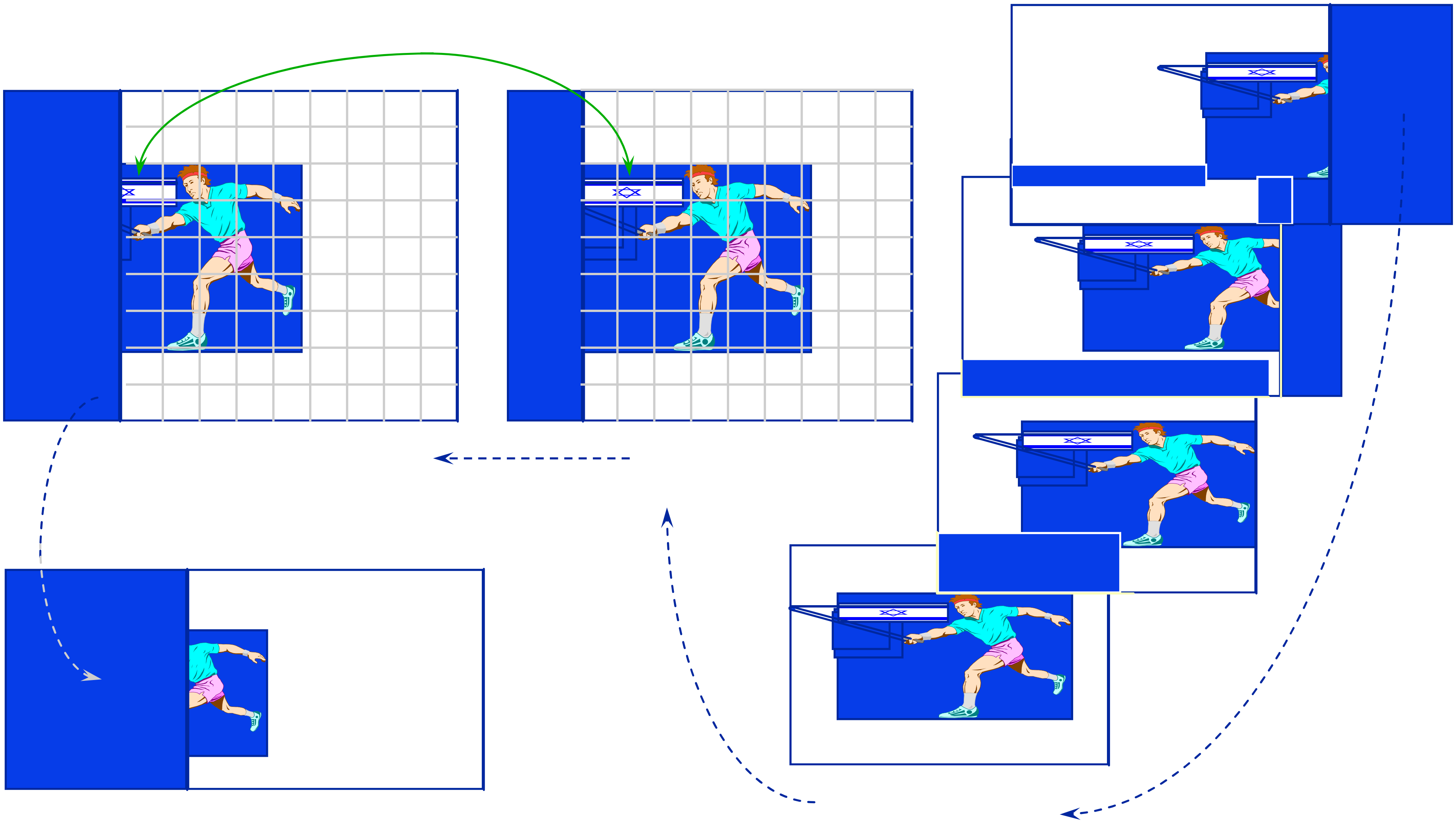
- Pixels in changed region are encoded after motion compensation process

Motion Estimation

- Estimates the displacement of image structures from one frame to another, not necessarily true motion
- Coding: Difference and Motion Vector (MV)



The Technique: Block Matching



Motion Detection

- Amount of changed data varies from frame to frame (VLC): a *buffer* is needed
- A *motion detection* algorithm should be used
- MC-DPCM is efficient if displacement vectors estimation is accurate !

Motion Compensation (Cont'd)

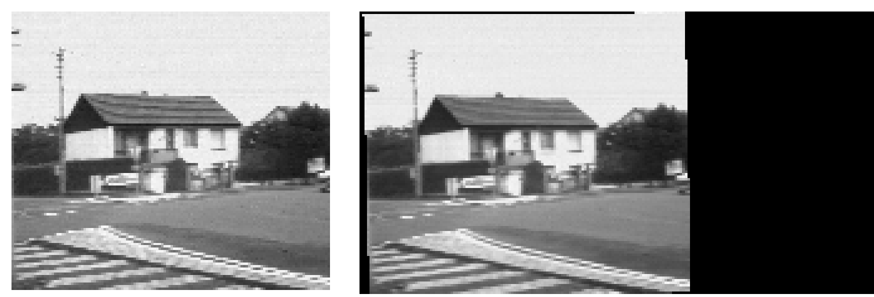
- Ideally: motion info for each pixel
  - too expensive
- Semantically: motion data for each region or object
  - second generation coding techniques
- Simplified: motion info
- for each 16x16 macroblock

Motion Model

- Affine motion:
  - translation, rotation, scale
  - 6 parameters in 2D
  - complex

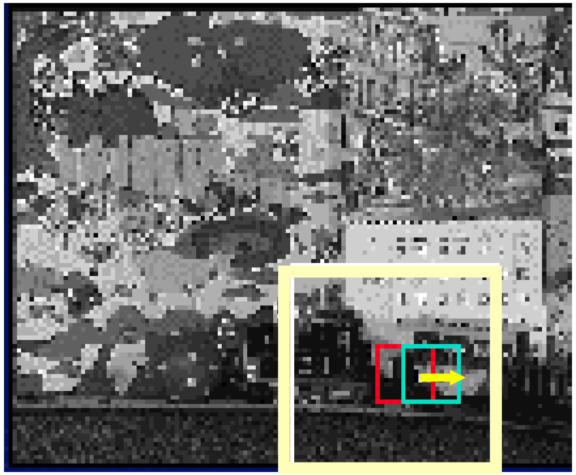
- Simplified translational model:
  - motion vectors (MV):

  2 parameters



ME Options

- optimal motion vector?
  - investigate all positions within a search window
  - keep the one with minimum Mean Square Error
  - MV = corresponding translation



Motion Estimation Techniques

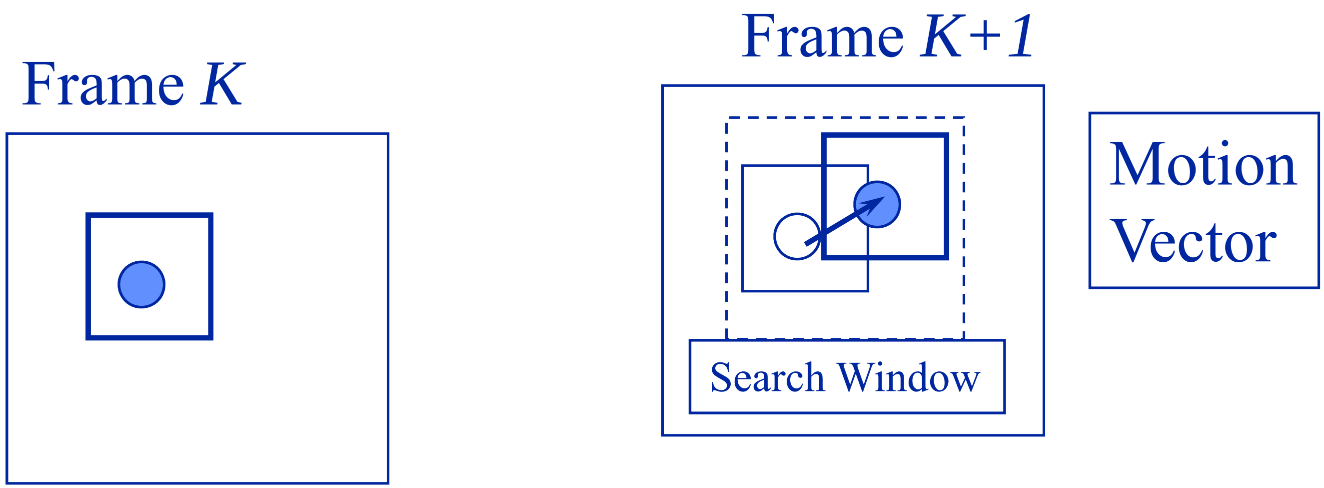
The choice of sending zero MV, is always available, with no cost…and worst results.

- Backward prediction
  - Predict where the pixels in a current frame were in a past frame
- Forward prediction
  - Predict where the pixels in a current frame will go to in a future frame

Motion Estimation

Block Matching Method

- Motion vector is estimated by pixel domain search procedure
- Most popular due to less h/w complexity
- Basic idea:



Matching Criteria

- Maximum cross-correlation
- Maximum pel matching count (MPC)
- Minimum mean squared error (MSE)

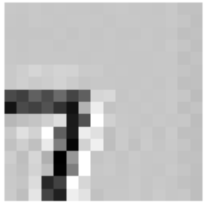
  Not popular in VLSI due to square operation
- Sum of absolute difference (SAD)

$$SAD = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \left| curr\_block(i, j) - ref\_region(i, j) \right|$$

Very popular but problematic due to several possible local minima

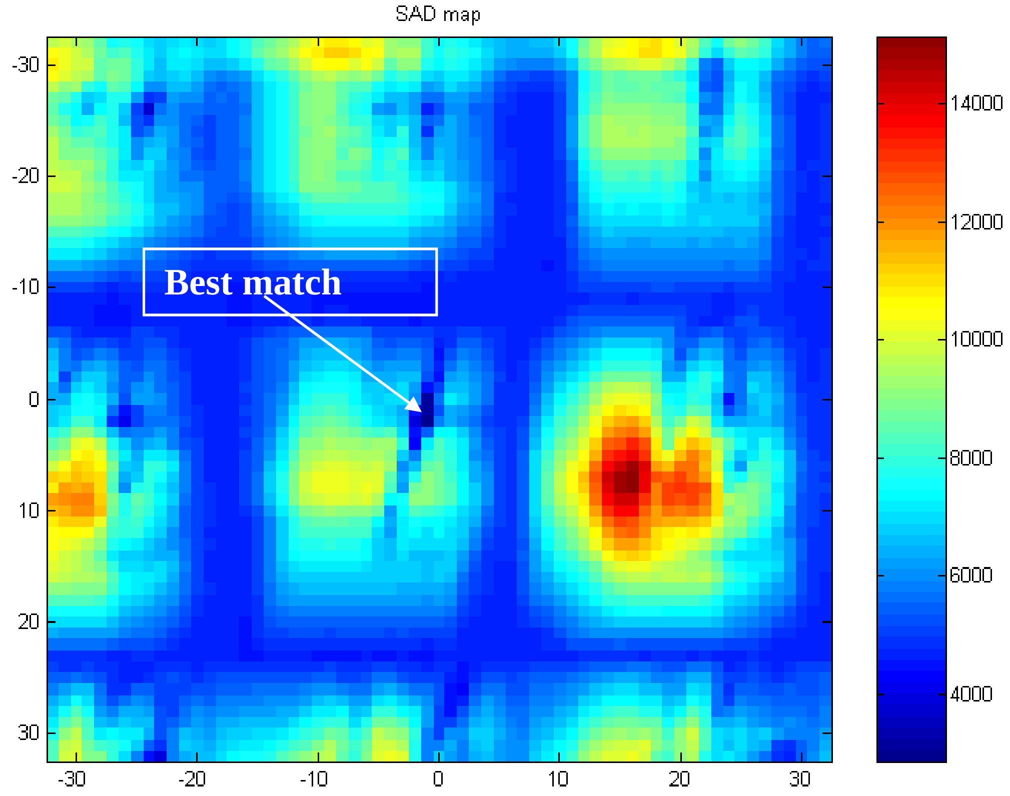
Motion Estimation - Example

Current block



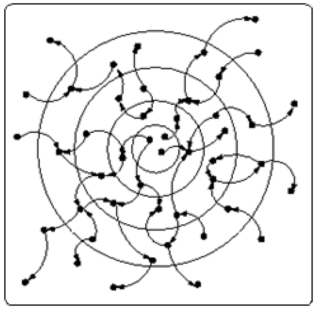
Reference region

SAD map



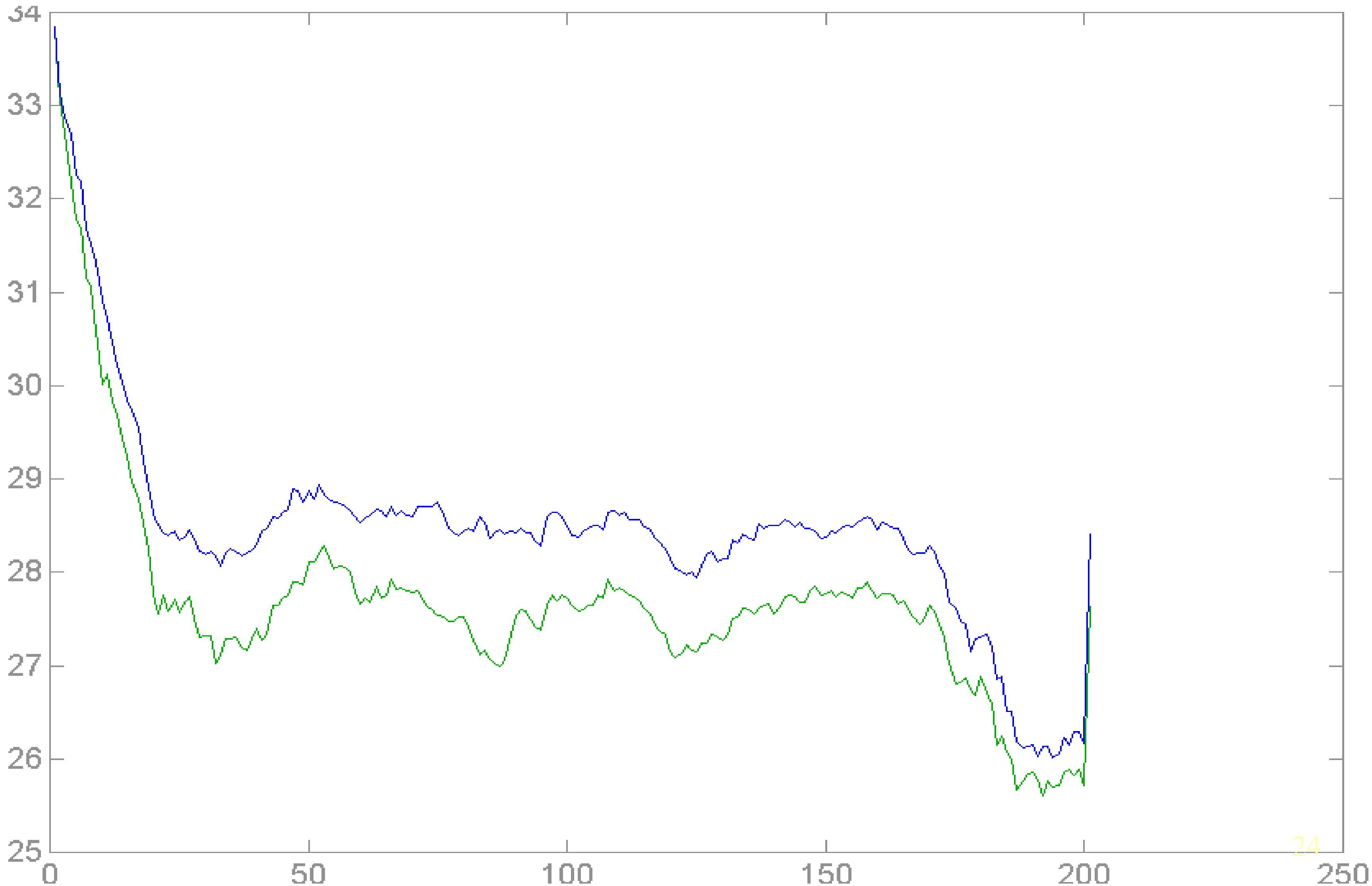
Other Parameters

- Complexity and implementation



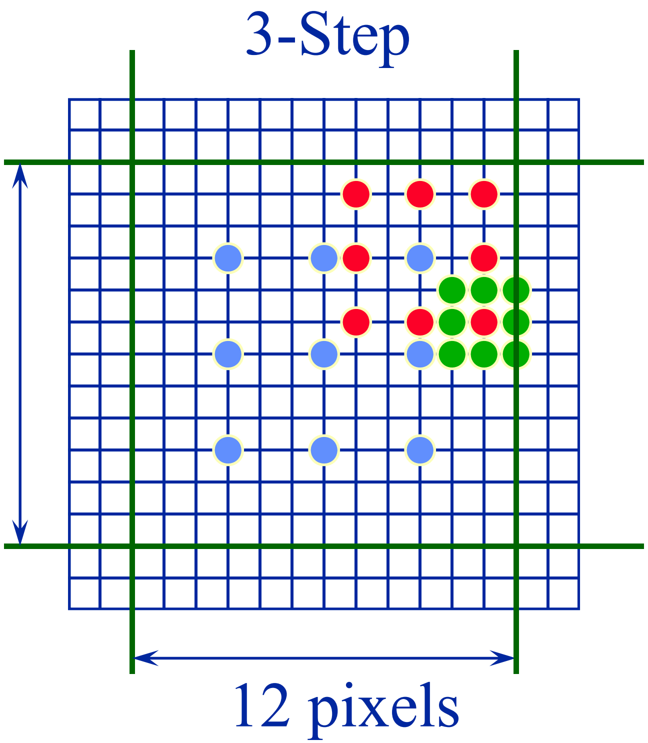
- Visual results ...

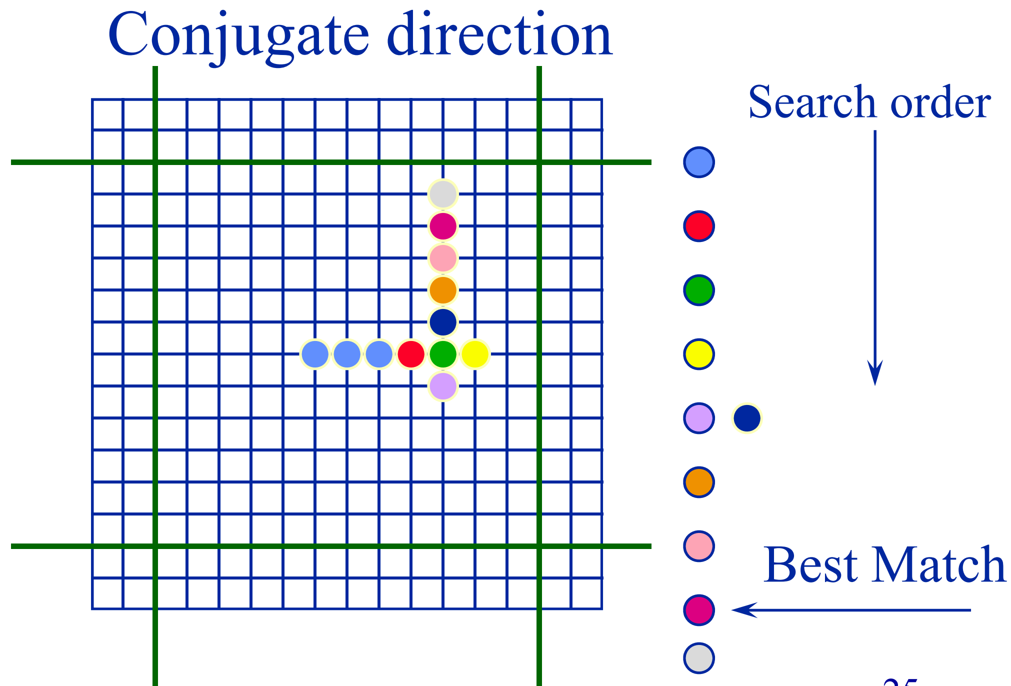
Full Search & Zero MV results



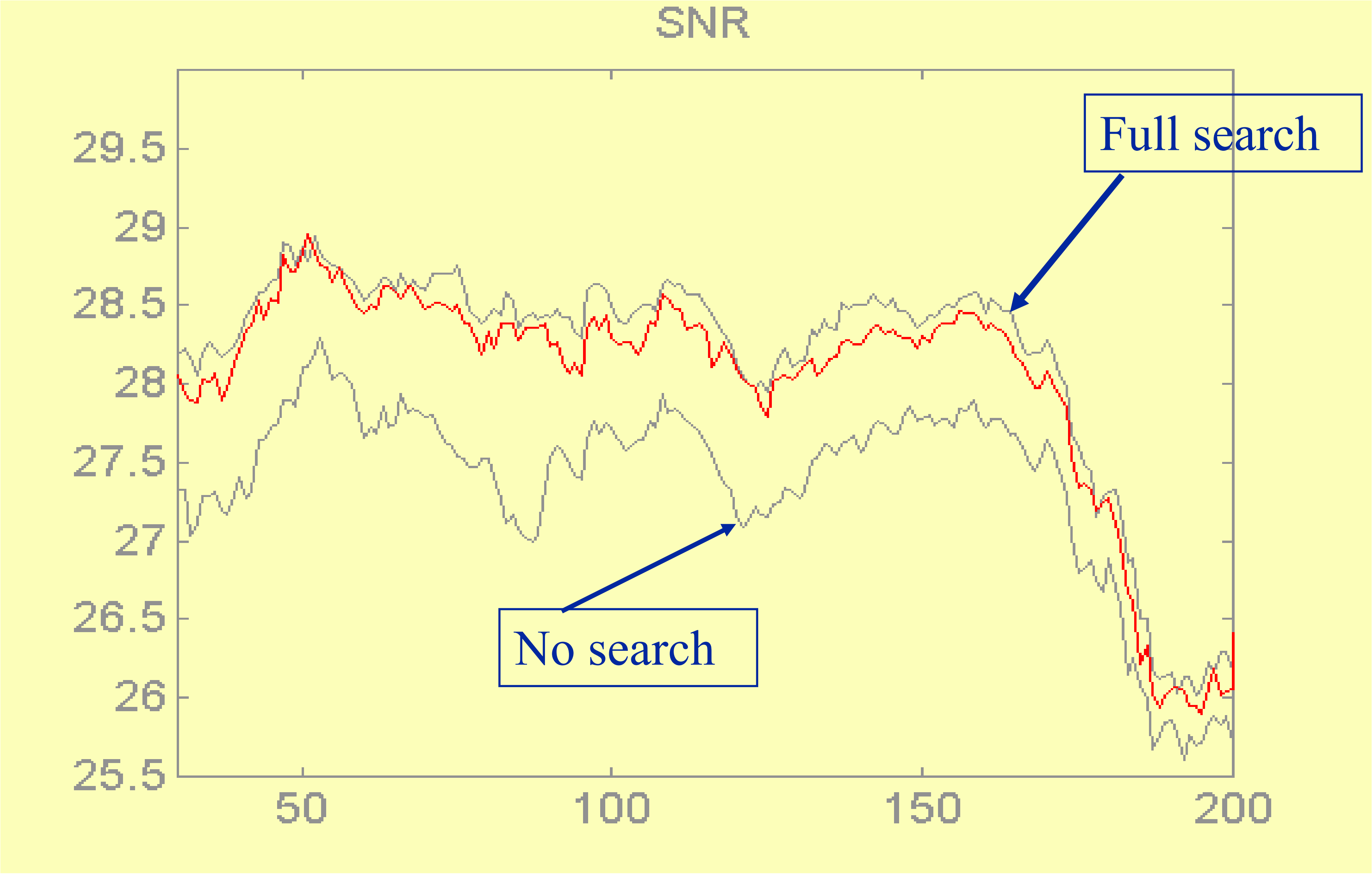
Search Procedure

- Full search (inside the search window)
- Three step
- Conjugate direction

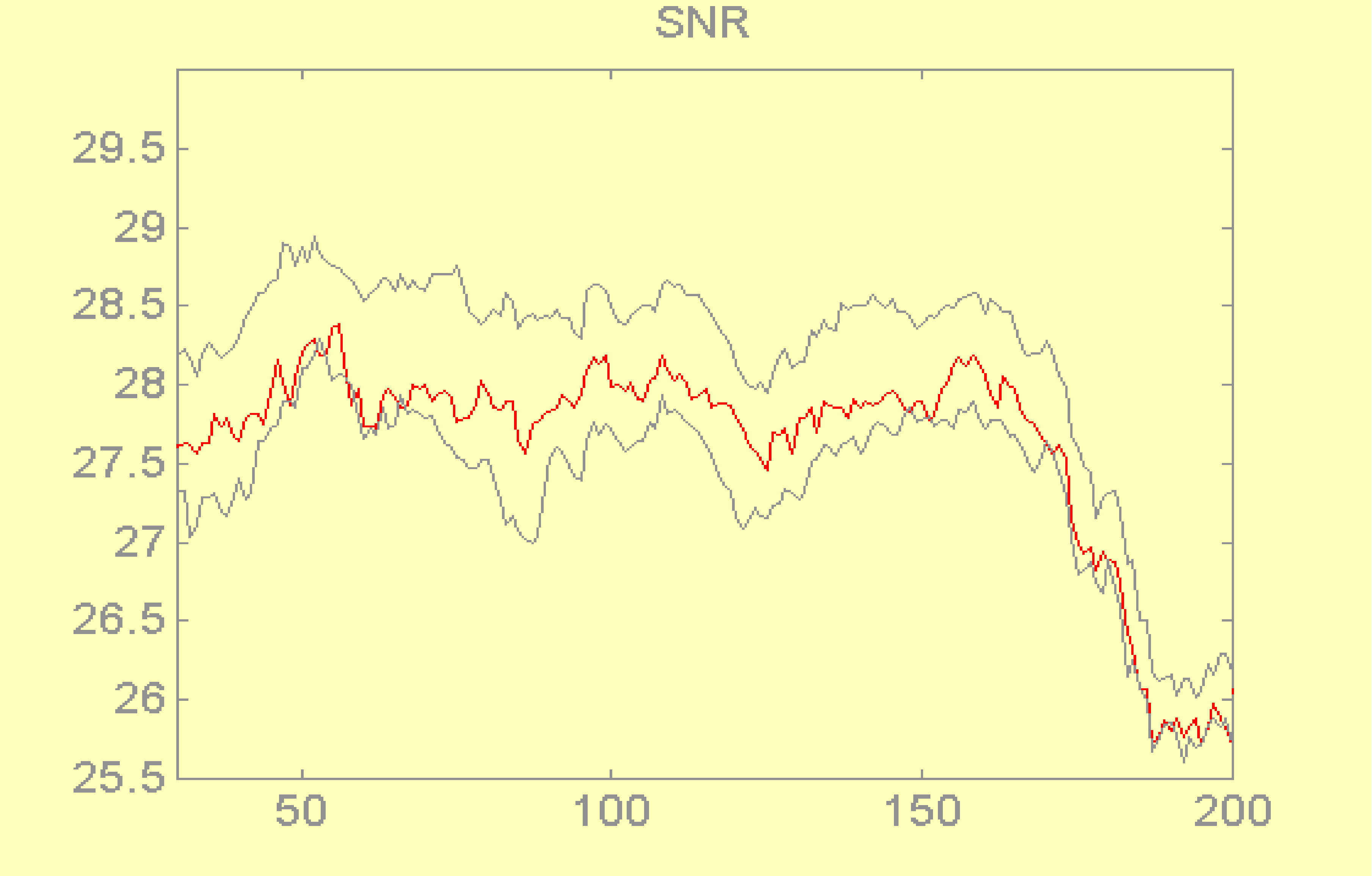




3 or 4 Step Search Results



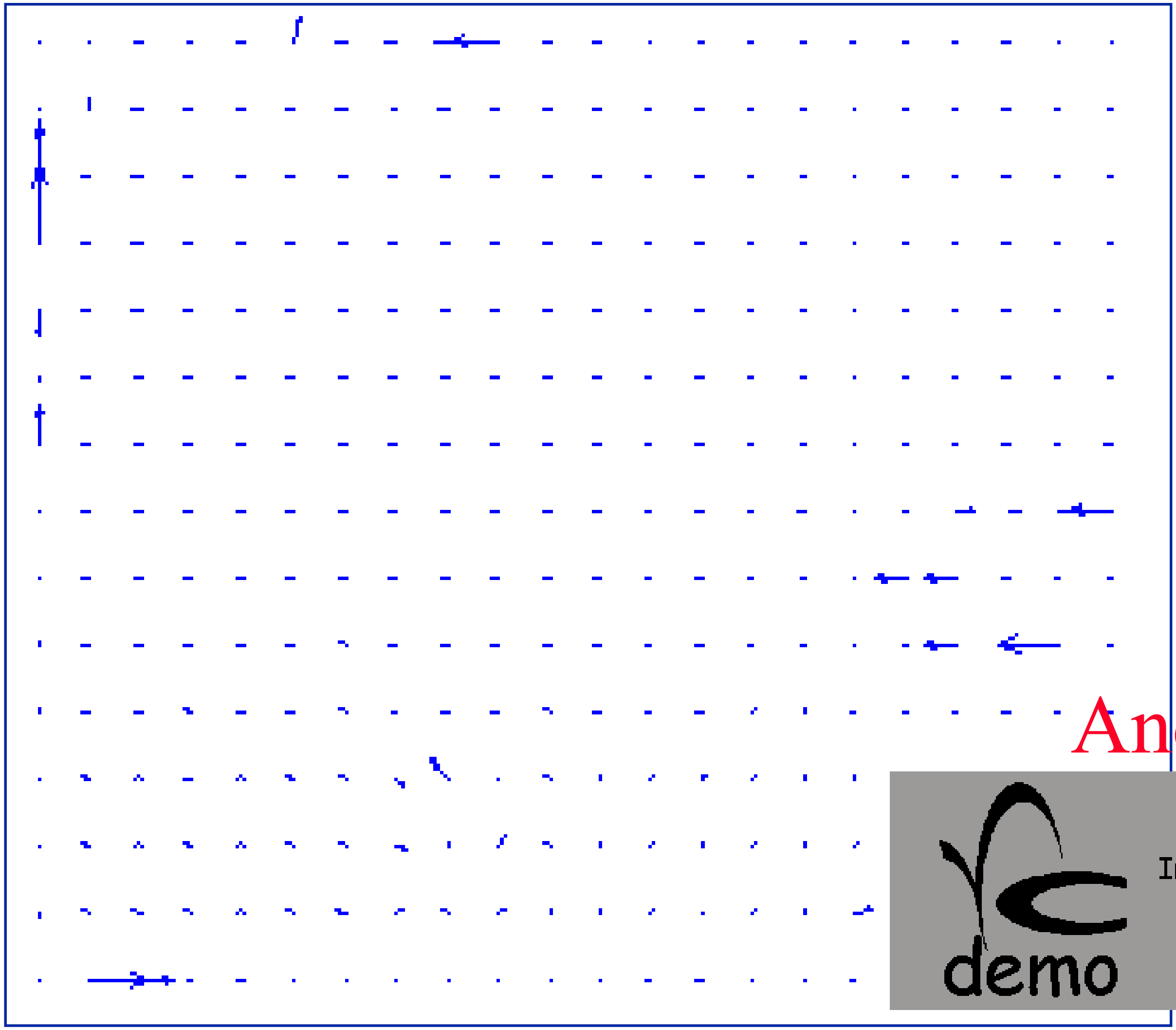
Conjugate Direction Search



Search Algorithm results

Algorithm	Ave. SNR(dB)	Caculations
Full search	28.46	100% (~7M)
Zero Search	27.7	0
Conjugate	27.9	2%
Orthogonal	28.1	3.6%
Decimation	28.2	33%
Pyramid and more...		

Motion Vectors Map - Example

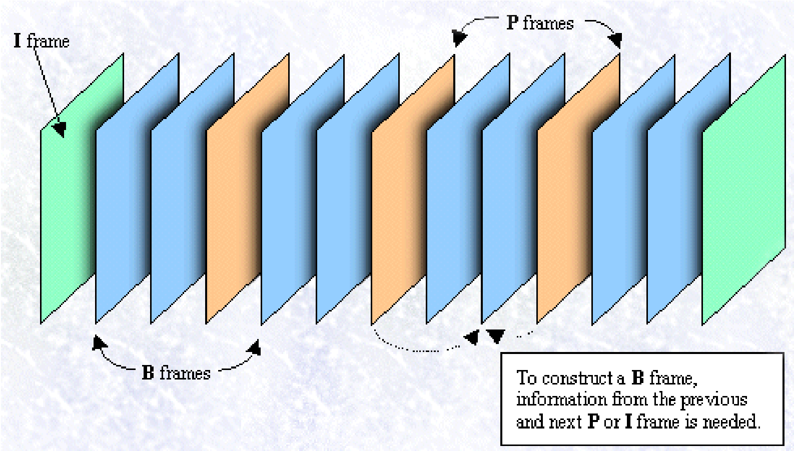


And now….

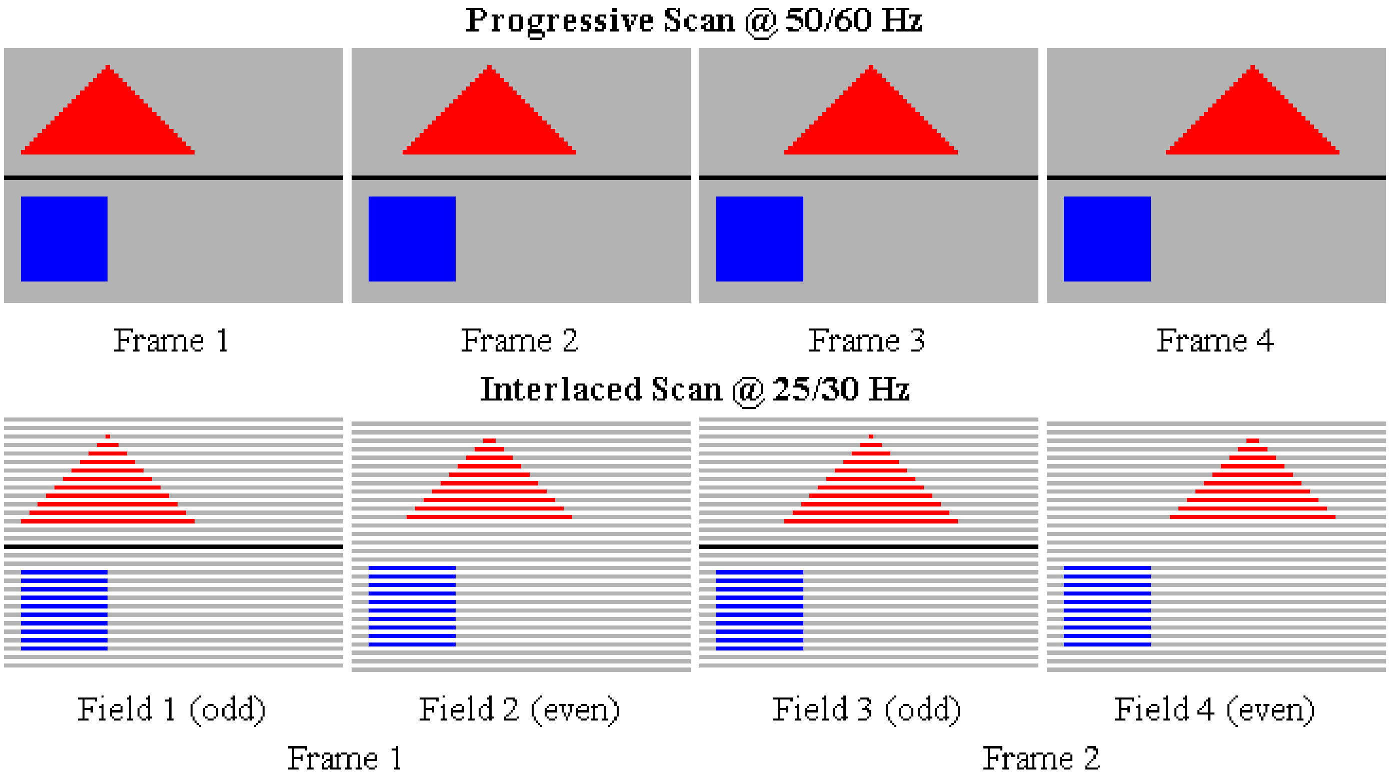
MC Transform Coding

- Temporal prediction error is 2D transform coded by segmenting the displaced frame difference into blocks, and encoding the DCT coeff. of each block

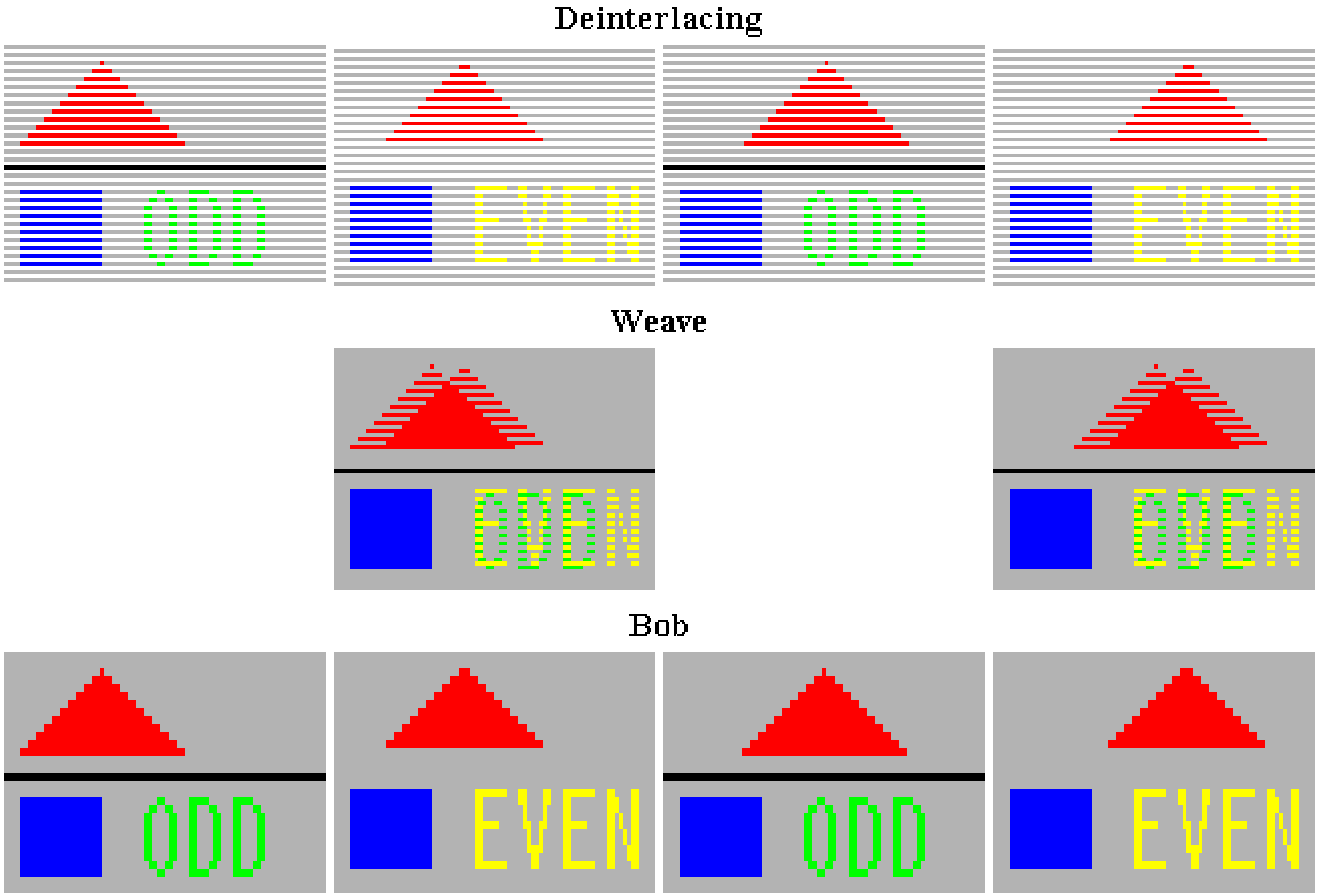
- *Intra* and *Inter* modes



Field and frame modes



Video DeInterlacing



Other MC Compression Techniques

- Vector Quantization:

  Prediction error signal encoded by VQ

- Sub-Band Coding:

  Frame prediction error (residual) is decomposed into 2D subbands

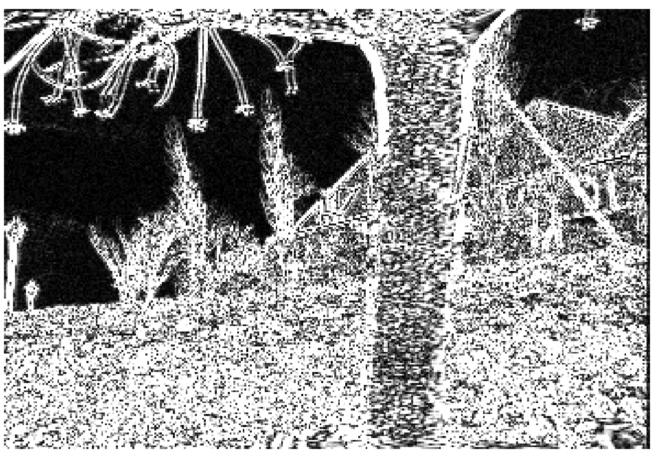
Video Compression (Cont'd)

- Complete process:
  - Encode every N frame "JPEG style"
  - Between the "JPEG style" frames, predict the motion from frame to frame
  - Subtract the predicted frame from the original one and encode the difference
- **Compression ratio: about 100:1 and more**

Motion Compensation Example

•frame #0

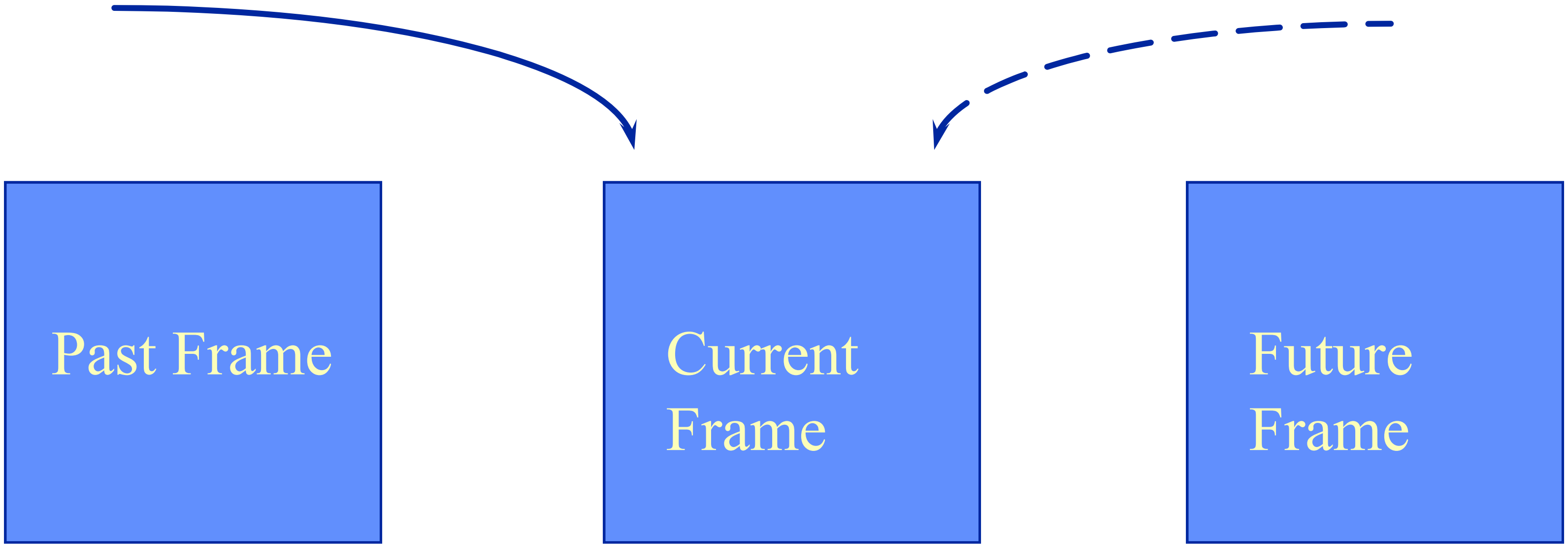
•frame #1



•difference frame

motion compensated difference frame

Prediction Technique



Bidirectional Interpolation

Why Bi-directional reference ?