Digital Video Processing

Introduction

Partially Adopted from New Mexico Tech and IITM

Digital Video Processing

- Digital Video
  - Digitized.
  - A sequence of images along the temporal axis.
- Processing
  - A running program or computing operation.
  - Driven by the real-world applications (e.g., segmentation, compression, filtering, retrieval).

Digital Video Applications

- Surveillance & Security
- Event Detection
- Protect copyright
- Home and health care
- Medical Video
- Entertainment
- Traffic and crowd monitoring
- Missing object detection
- Autonomous vehicle guidance
- Transmission and playback

A Brief History

- Birth of Television (1920s)
- Cable TV system (1968)
- Video games (1970s)
- All-digital HDTV (1990s)
- Video streaming (2000s)
- Everyday video transmission through internet and wireless networks

Why Video?

- The magic of Tele-Vision
  - Our vision capability is extended in **space**.

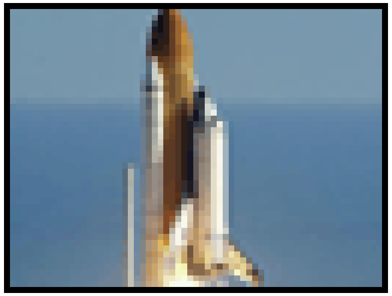
You don’t need to travel to north pole to watch polar bears.

Why Video?

- Our vision capability is extended in **time**.
  - If time can be reversed, we will not need a Gigabyte hard-drive to store the moments of how a baby is growing.

- The fundamental interplay between **time** and **motion**.
  - We measure time by the motion of material things.
  - Motion offers a new horizon for us to understand the world.

Diversity of Motion



Motion in Video

- It is not an arbitrary concatenation of images, but a sequence of images carrying a *coherent* interpretation of natural scene.
  - Ordering is important.
  - Sampling rate is important.
  - The role of a single frame is less important due to the masking effect of Human Visual System (HVS).

Classification of Video Coding Systems

- Analogue (signals predominantly in analogue form)
  - NTSC - 6MHz (4.2MHz video), 29.97 fps
  - PAL - 6-8MHz (4.2-6MHz video), 25 fps

- Color spaces
  - RGB, YUV, YC_bC_r, YIQ

Analog Signals

- Common consumer display mechanism for video, the television, still uses analogue display devices such as CRT.

- Until all terrestrial and satellite broadcasts become digital, analogue video formats will remain significant.

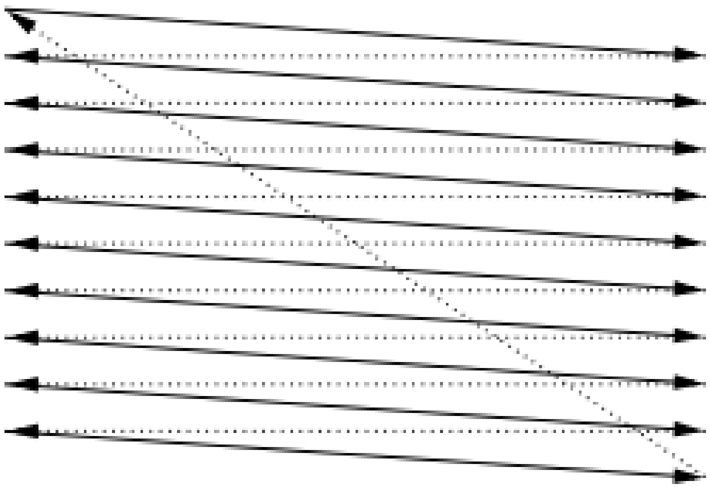
- Principal formats
  - NTSC (National Television Systems Committee)
  - PAL (Phase Alternate Line)
  - SECAM (Sequential Colour with Memory)

- The formats take advantage of the persistence of human vision by using interlaced scanning pattern.
  - odd and even lines of each picture are read out in two separate scans of the odd and even fields respectively.

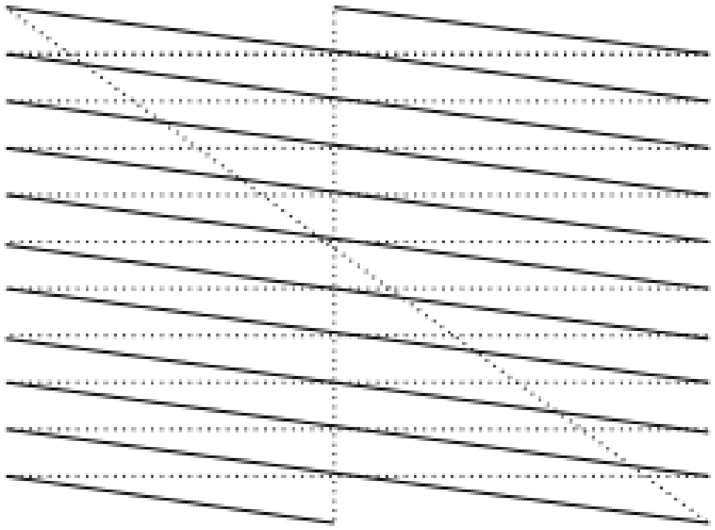
Progressive and Interlaced Scan Pattern

- Progressive scan patterns are used for high resolution displays like computer CRT monitors, digital cinema projections.
- In progressive scan, each frame of picture information is scanned completely to create the video signal.
- In interlaced scan pattern, the odd and even lines of each picture are read out in two separate scans of the odd and even fields respectively.

Progressive and Interlaced Scan Pattern



Progressive scanning



Interlaced Scanning

Scan Pattern Problem

- Consider the following two raster scan formats:
  - progressive scan using 20 frames/second, 500 lines/frame, and
  - interlaced scan using 40 fields/second, 250 lines/field.
- For each scan format, determine
  i. The overall line rate
  ii. The maximum temporal frequency of the signal.
  iii. The maximum vertical frequency of the signal.

Soln.: Progressive scan

i. The overall line rate is $20 \times 500 = 10000$ lines/sec
ii. The maximum temporal frequency (when successive frames alternate between black and white) is half of the frame rate $= 10$ cycles/second.
iii. The maximum vertical frequency (when successive lines alternate between black and white) is half of the line numbers per frame $= 250$ cycles/picture-height.

Scan Pattern Problem

- Consider the following two raster scan formats:
  - progressive scan using 20 frames/second, 500 lines/frame, and
  - interlaced scan using 40 fields/second, 250 lines/field.
- For each scan format, determine
  - i. The overall line rate
  - ii. The maximum temporal frequency of the signal.
  - iii. The maximum vertical frequency of the signal.

Soln.: Interlaced scan

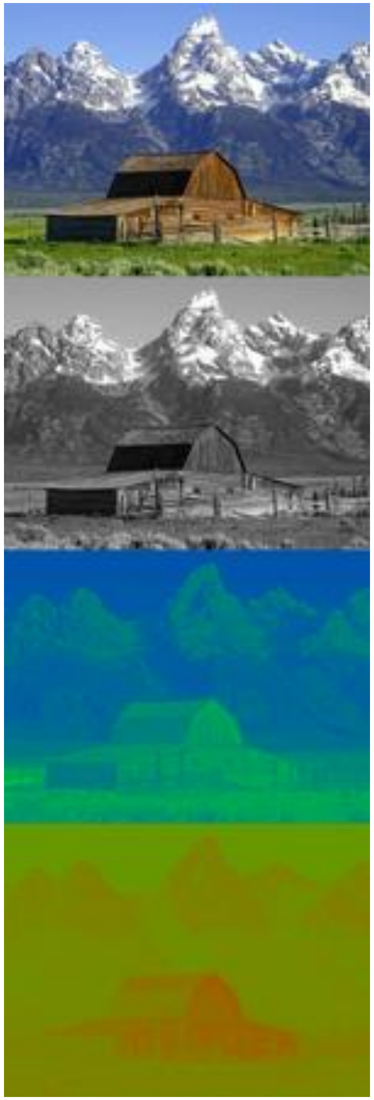
i. The overall line rate is $40 \times 250 = 10000$ lines/sec

ii. The maximum temporal frequency (when successive fields alternate between black and white) is half of the field rate =20 cycles/second.

iii. The maximum vertical frequency (when successive lines alternate between black and white) is half of the line numbers per field = 125 cycles/picture-height.

Colour Coding: YUV

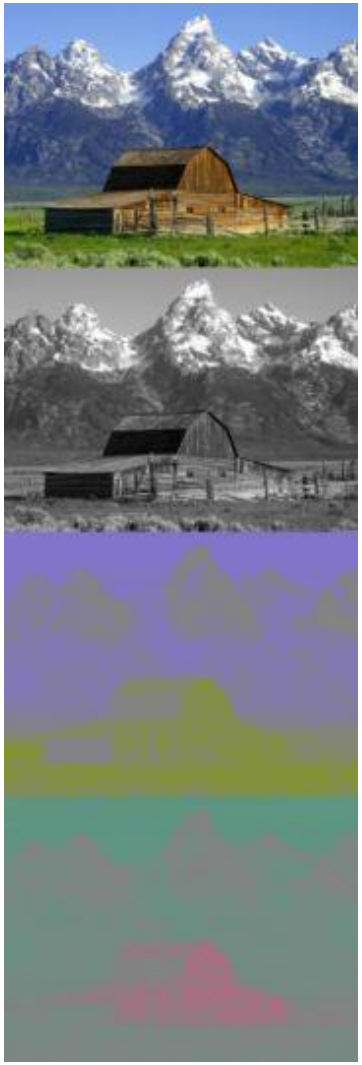
- PAL video standard
  - Y is luminance (brightness)
  - UV are chrominance (color)
- YUV from RGB
  - $Y = .299R + .587G + .114B$
  - $U = 0.492\ (B - Y)$
  - $V = 0.877\ (R - Y)$



An image along with its Y, U, and V components respectively

Colour Coding: YC_bC_r

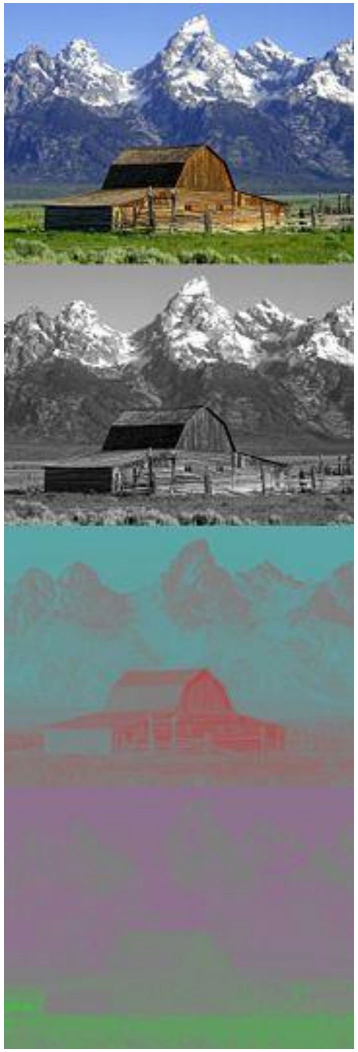
- Subset of YUV that scales and shifts the chrominance values into range 0…1.
  - $Y = 0.299R + 0.587G + 0.114B$
  - $C_b = ((R-Y)/1.6) + 0.5$
  - $C_r = ((B-Y)/2) + 0.5$



A colour image and its Y, C_b and C_r components. The Y image is essentially a grayscale copy of the main image.

Colour Coding: YIQ

- NTSC standard
- YIQ from RGB
  - $Y = .299R + .587G + .114B$
  - $I = .74\ (R - Y) - .27\ (B - Y)$
  - $Q = 0.48\ (R - Y) + 0.41\ (B - Y)$



An image along with its Y, I, and Q components.

Digital Video

- In a digital video, the picture information is digitized both spatially and temporally and the resultant pixel intensities are quantized.

Block diagram depicting the process of obtaining digital video from continuous natural scene

Video Formats

- Digital video consists of video frames that are displayed at a prescribed frame rate.
  - A frame rate of 30 frames/sec is used in NTSC video.
- The frame format specifies the size of individual frames in terms of pixels.

Format	Luminance Pixel Resolution	Typical Applications
Sub-QCIF	128 X 96	Mobile Multimedia
QCIF	176 X 144	Video conferencing and Mobile Mulimedia
CIF	352 X 288	Video conferencing
4CIF	704 X 576	SDTV and DVD-Video
16CIF	1408 X 1152	HDTV and DVD-Video

Spatial Sampling

- The sensitivity of Human Visual System (HVS) varies according to the spatial frequency of an image.
- In the digital representation of the image, the value of each pixel needs to be quantized using some finite precision.
- In practice, 8 bits are used per luminance sample.

Temporal Sampling

- A video consists of a sequence of images, displayed in rapid succession, to give an illusion of continuous motion.

- If the time gap between successive frames is too large, the viewer will observe jerky motion.

- The sensitivity of HVS drops off significantly at high frame rates.

- In practice, most video formats use temporal sampling rates of 24 frames per second and above.

Video Acquisition

Video camera

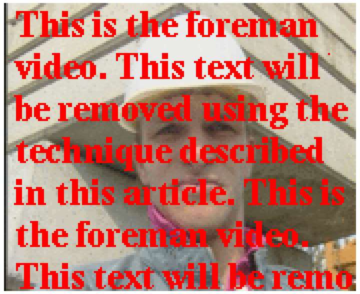
VHS digitization

computer-generated

Video Manipulation

- Why?
  - Fight against a non-ideal video acquisition (e.g., analog heritage, film scratches, limited resolution) or transmission environment.
  - Create new and artificial video content (e.g., spatio-temporal interpolation, background/foreground modification).

Video Inpainting



Application: Remove the annoying texts added by various video conversion software.

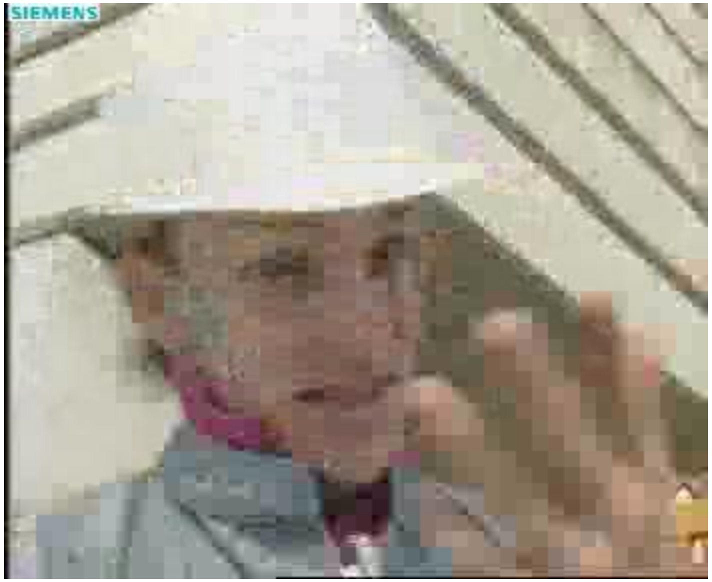
Error Concealment

Some blocks are corrupted
due to channel errors

Corrupted blocks are recovered
from surrounding neighbors
in space and time

Post-processing

Deblocking: suppress block artifacts in video



Decoded video frame
at very low bit rate

Processed video frame
after deblocking

Video Matting

Motion Estimation

Features of Moving Pictures

- Moving images contain significant temporal redundancy.
  - successive frames are very similar.

Frames captured at 1/10 second intervals

Motion Estimation and Compensation

- The amount of data to be coded can be reduced significantly if the previous frame is subtracted from the current frame.

Residual Image

Motion Estimation

Motion estimation is the process of determining motion vectors that describe the transformation from one 2D image to another; usually from adjacent frames in a video sequence.

It is an ill-posed problem as the motion is in three dimensions but the images are a projection of the 3D scene onto a 2D plane.

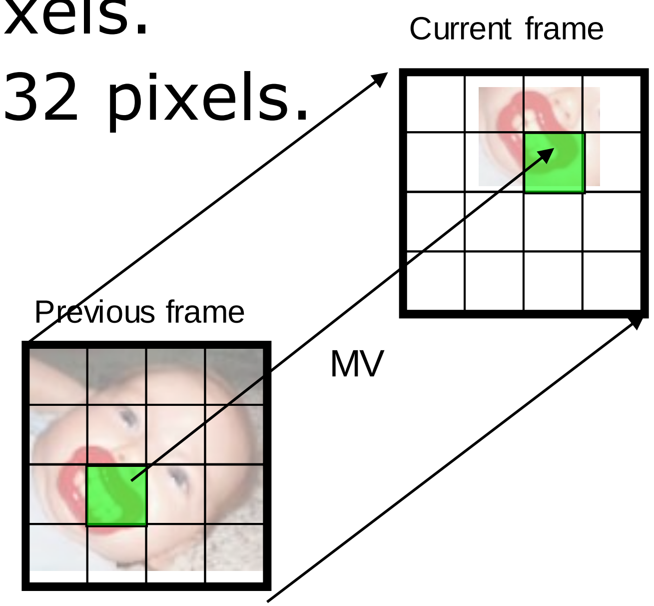
Block-Matching

- The objective of block-matching is to find a candidate block in the search region best matched to the source block.

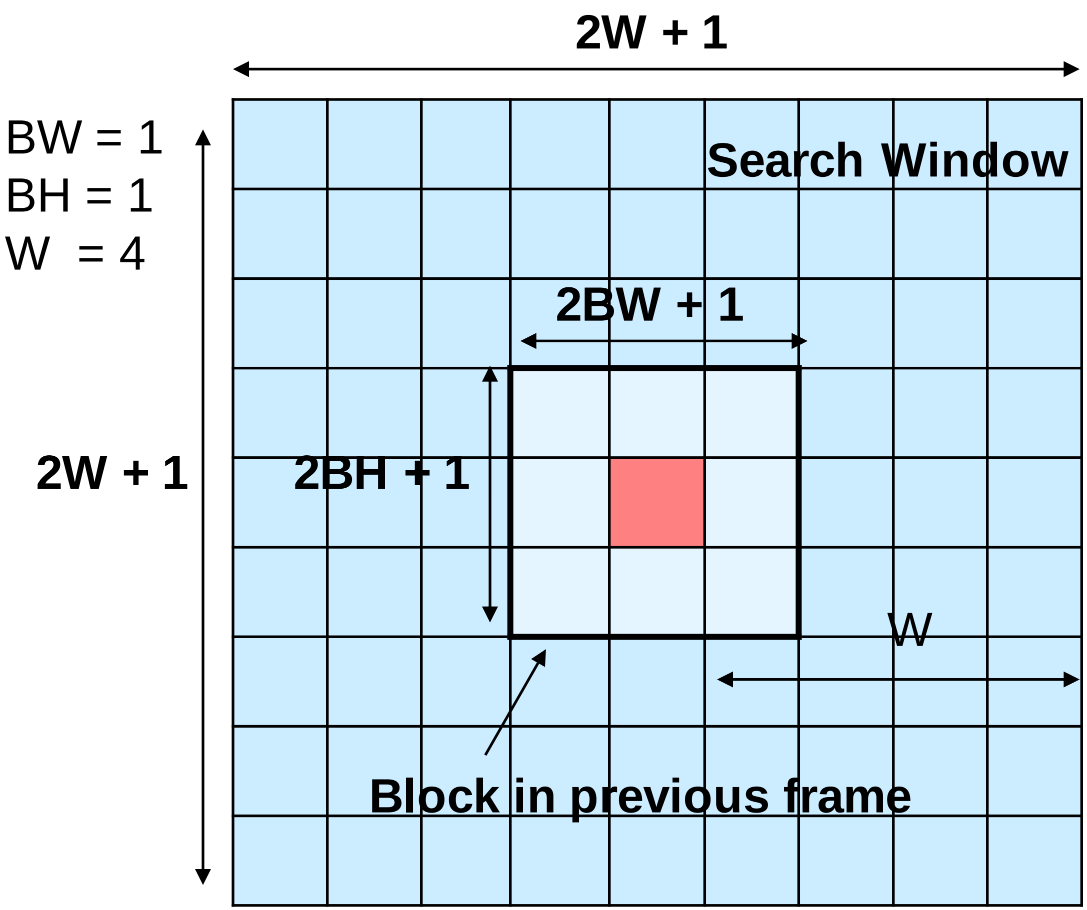
- The relative distances between a source block and its candidate blocks are called motion vectors.

Block-Matching

- Divide the previous frame into small rectangular blocks.
- Find the best match for the reference block in current frame.
- Calculate motion vector between the previous block and its counterpart in the current frame.
- Typical size for a block: 16x16 pixels.
- Search Range W: typically 16 or 32 pixels.
- Similarity Measures:
  - Mean Absolute Error (MAE)
  - Mean Square Error (MSE)



Block-Matching



Search Window
(in current frame)

A region which has the same center as the selected block in the previous frame, extended by W pixels in both directions

BH = Block Height
BW = Block Width

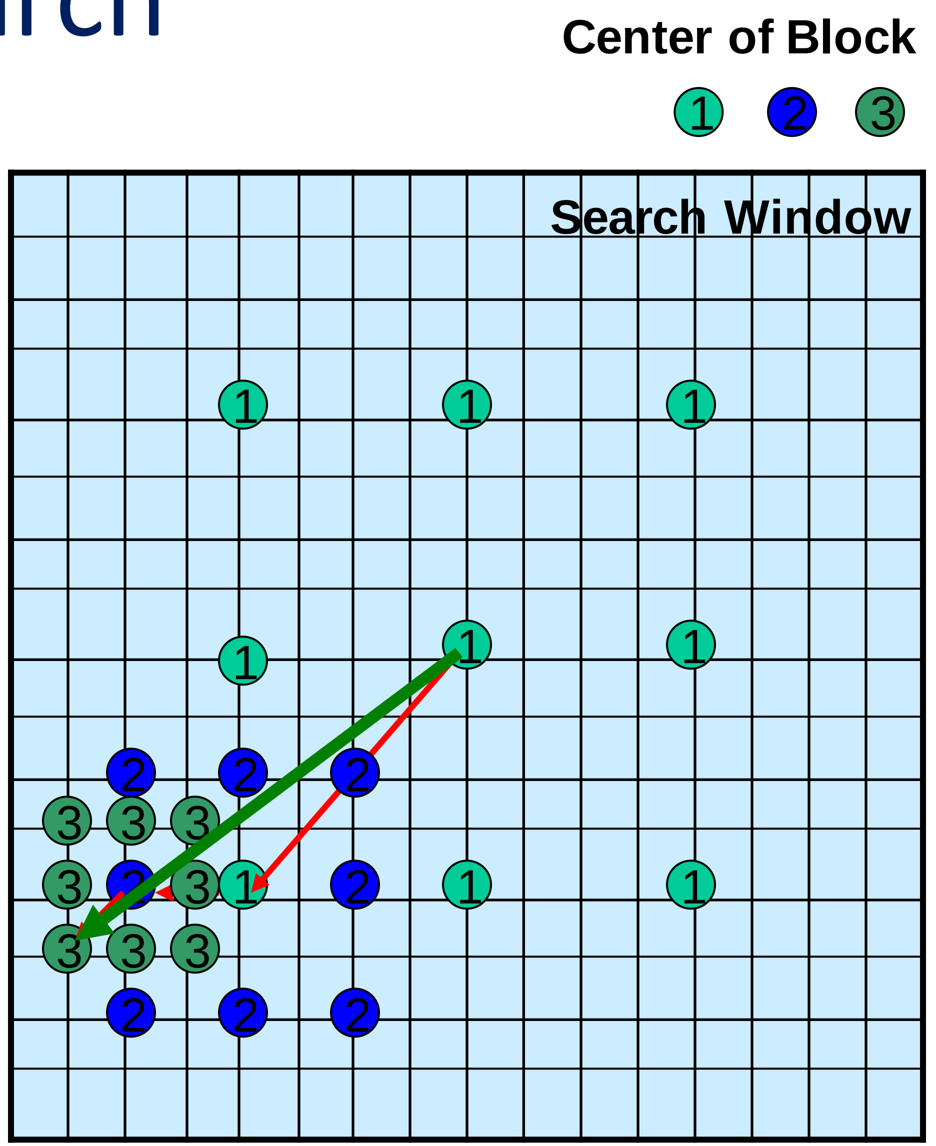
center pixel

Block-Matching

- Two kinds of methods commonly used:
  - Fast Search Algorithms
    - E.g.: 3-Step Search
  - Exhaustive Search Algorithm

Fast Search Algorithm: 3-Step Search

- 1st Step:
  - Search 8 surroundings and the central point.
  - Distance = w/2 pixels.
  - Find the best match.
- 2nd Step:
  - Use previous best match as center.
  - Repeat 1st step with distance = w/4 pixels.
- 3rd Step:
  - Repeat 1st step with distance = w/8 pixels.



3-Step Search

- Advantages:
  - Extremely fast.

- Disadvantages :
  - All fast algorithms greatly rely on a monotonically increasing match criteria around the location of the optimal motion vector.
    - Easily fall into local minimum.

  - Limited number of positions examined (only 25 points) inside the search window, only find suboptimal solution.

Exhaustive Search

- All candidates within search window are examined
- $(2W+1) \times 2$ positions should be examined
- Advantage: Good accuracy; Finds best match
- Disadvantage: High computational load. Impractical for real-time applications