Detection

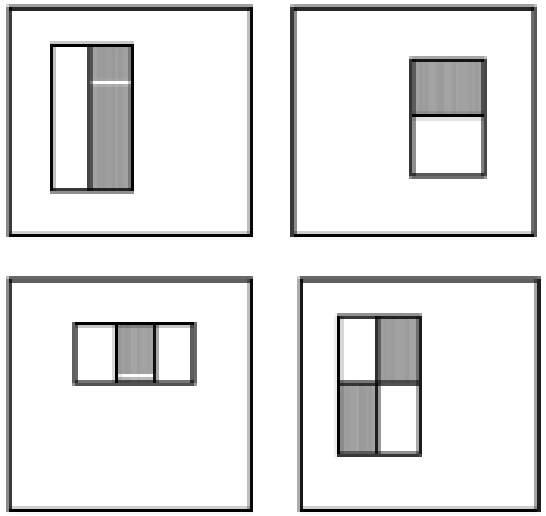
Sung Soo Hwang

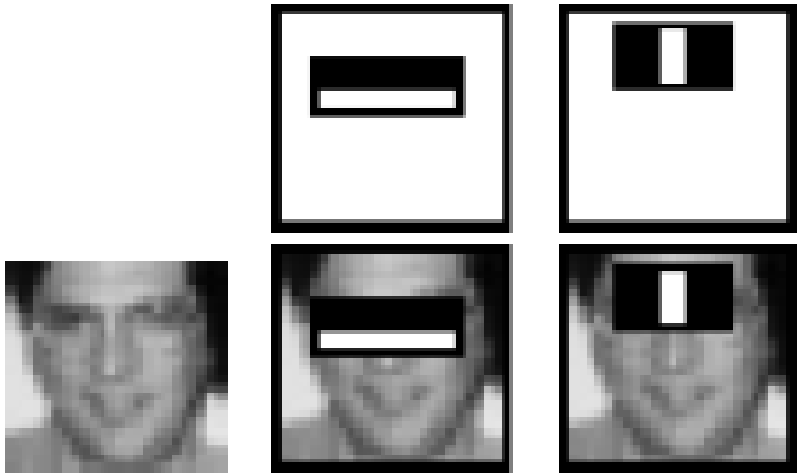
Introduction

- How can we make computers detect objects?
  - Training stage
    - Collect a large amount of object images and non-object images
    - Find features suitable to represent objects
    - Design a classifier(or threshold) to classify objects
  - Test stage
    - Extract features from the input image
    - Detect objects using the trained classifier

Face Detection

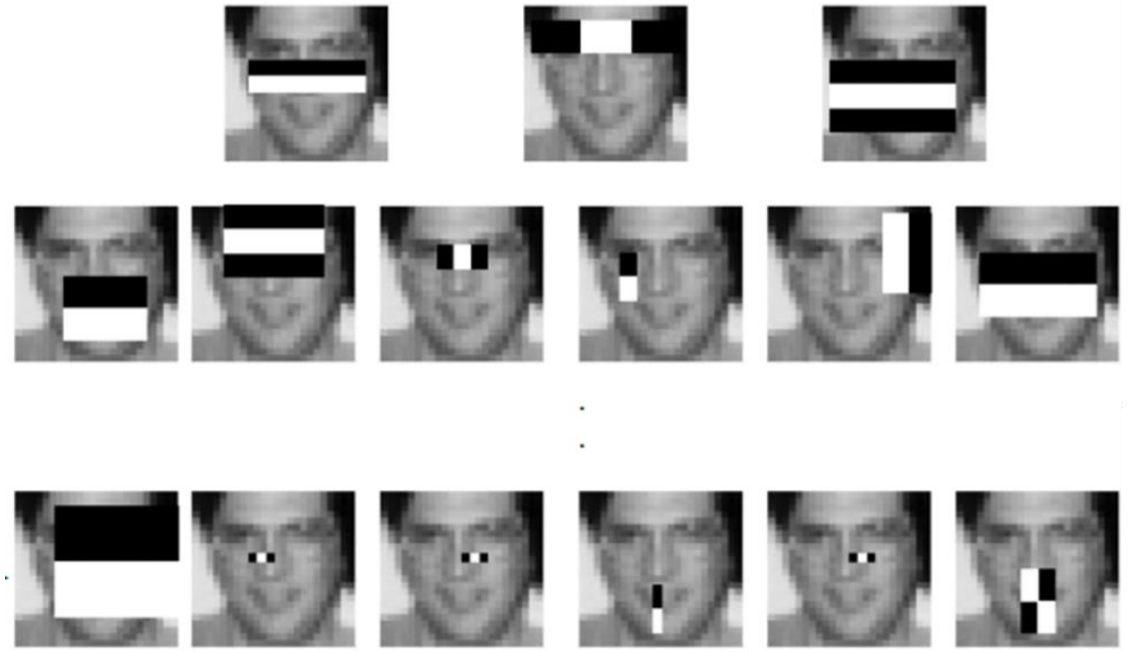
- Feature
  - Harr-like feature is used in openCV
    - It can be defined as the difference of the sum of pixels of areas inside the rec tangle
    - sum of pixels under black rectangle – sum of pixels under white rectangle
    - It can be at any position and scale within the original image





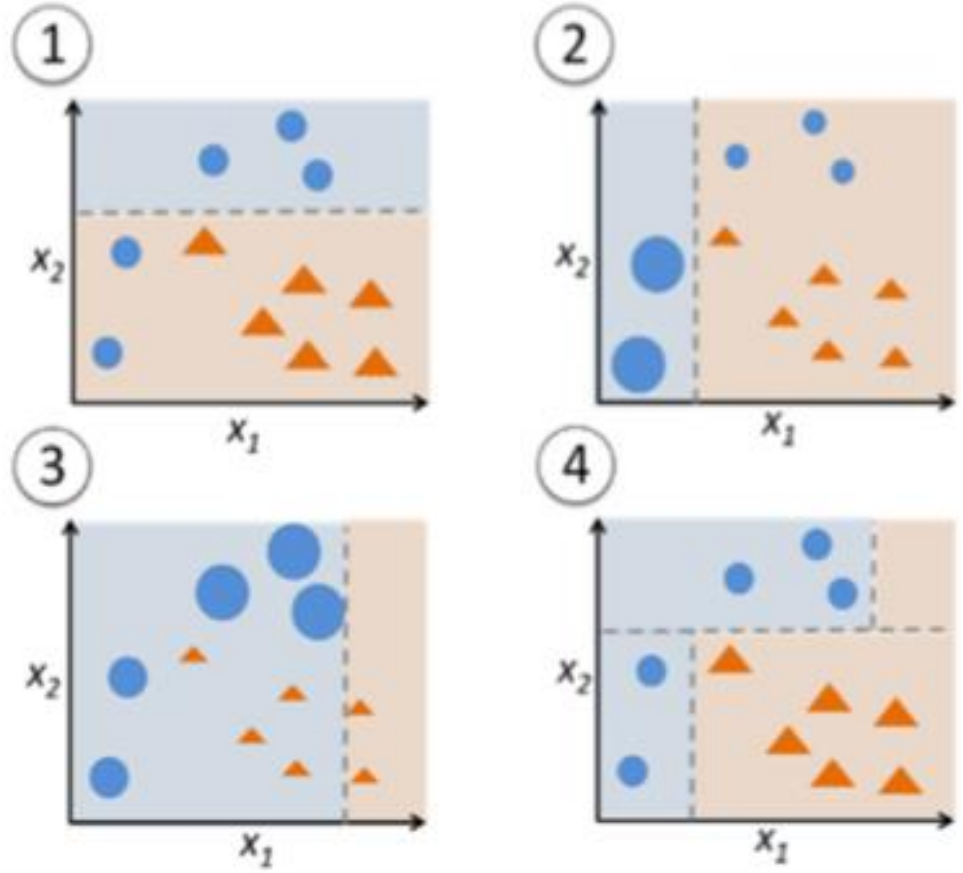
Face Detection

- Training
  - By changing size and location, lots of features can be generated
  - Among them, choose features that classify human faces



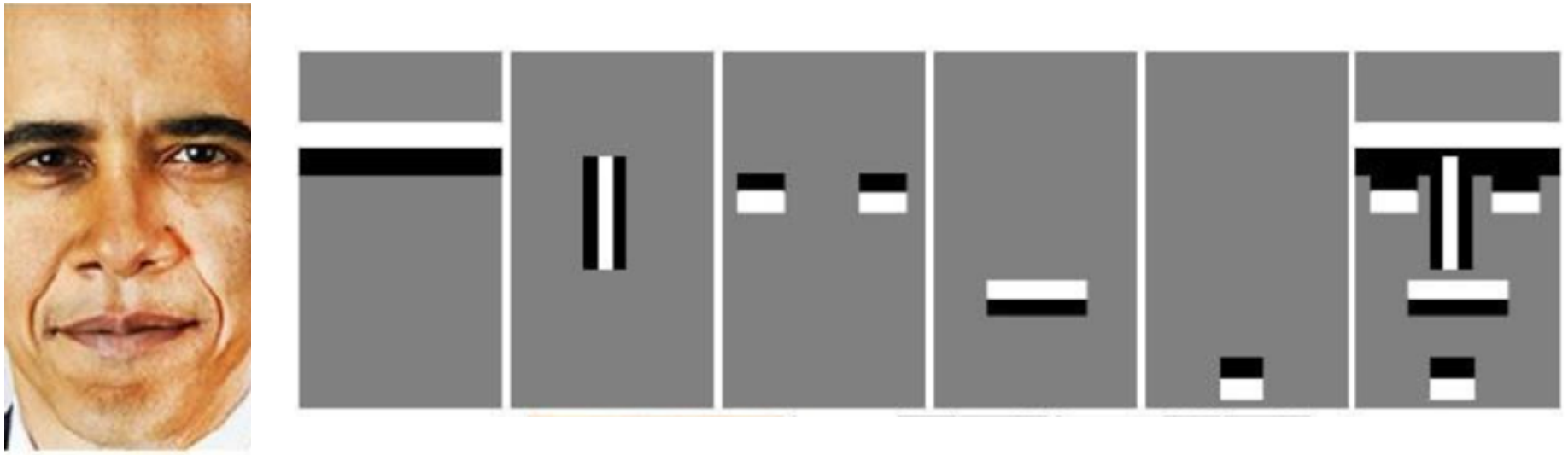
Face Detection

- Training
  - In openCV, Adaboost(Adaptive Boosting) is used
  - Boosting: A set of weak-learner generates a strong-learner
  - Adaptive: Weight of each sample is adjusted depending on the accuracy of already-trained weak learners



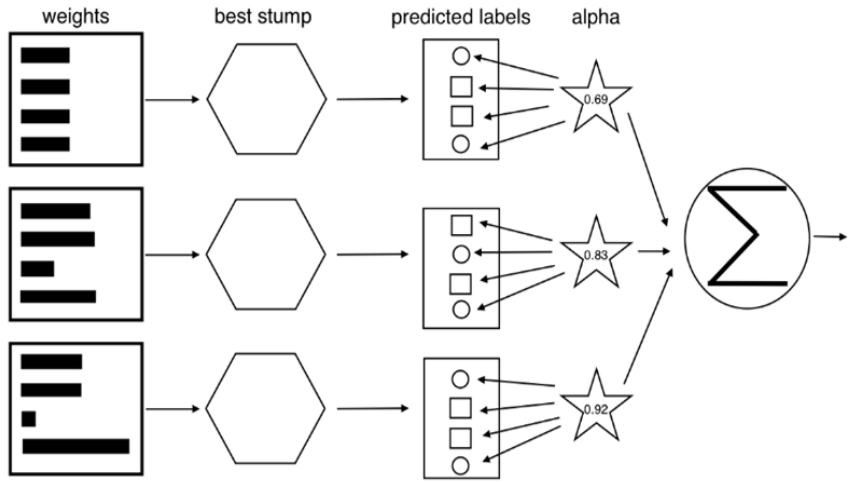
Face Detection

- Training
  - Face representation using Harr-like features

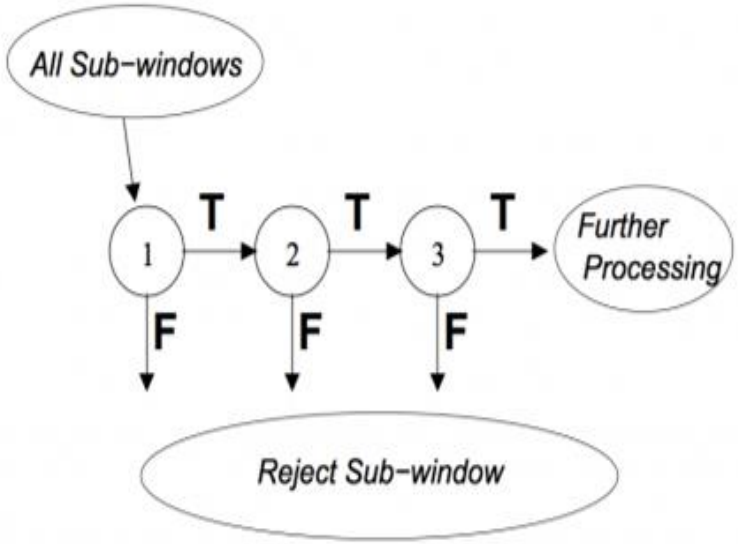


Face Detection

- Cascade classifier
  - Generate a strong learner using multiple weak learners



  - Each strong learner is connected in cascade
    - Number of weak learners in each strong learner: 3>2>1
    - Lots of non-face regions are easily eliminated



Face Detection

- Fast computation of Harr-like features(sum of pixels under a rectangle) can be conducted by using integral image

1	2	2	4	1
3	4	1	5	2
2	3	3	2	4
4	1	5	4	6
6	3	2	1	3

input image

0	0	0	0	0	0
0	1	3	5	9	10
0	4	10	13	22	25
0	6	15	21	32	39
0	10	20	31	46	59
0	16	29	42	58	74

integral image

unesco
University Network
Member

Ministry of
Education

Tracking

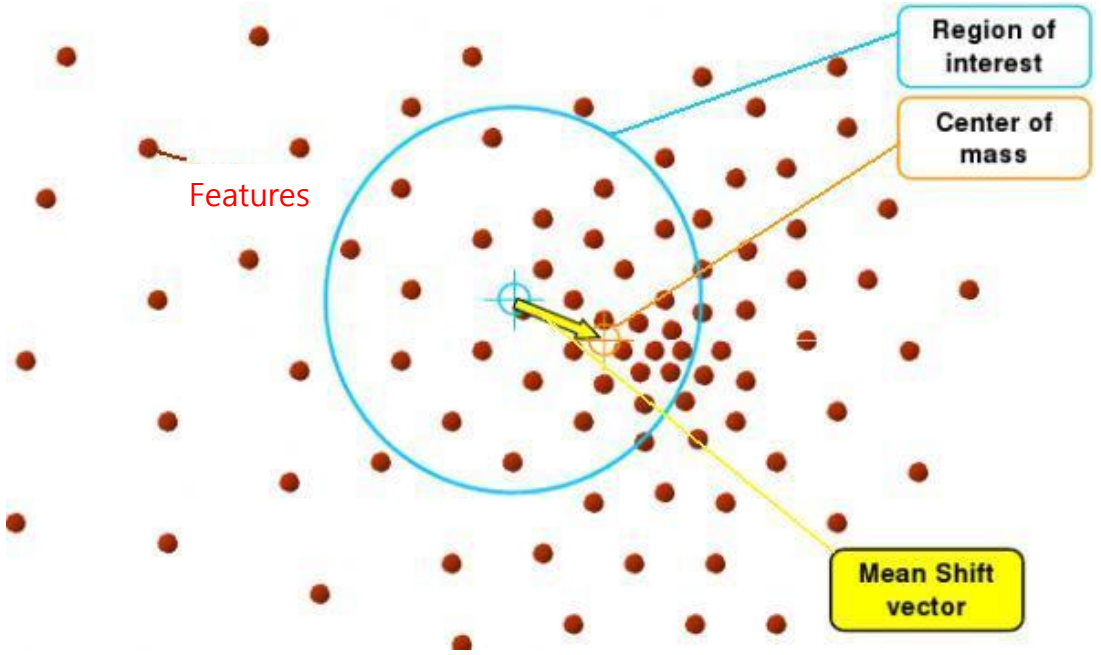
Sung Soo Hwang

Introduction

- Basic concept
  - First, a ROI is selected by user-interaction or detection
  - Represent the ROI with histograms or features
  - Find the best matching patch to the ROI at the next frame

Meanshift

- It is a procedure for locating the maxima of a density function given discrete data sampled from that function
- It is an iterative method

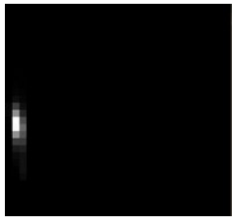


Meanshift

- Histogram back-projection

Model image

Hue-Saturation histogram



Back projection

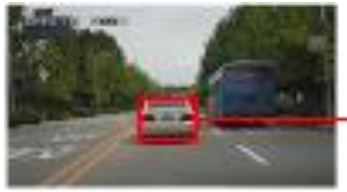
Target image

Probability of each pixel being part of model image

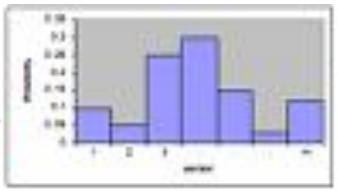
Meanshift

- Tracking using mean shift

initialization



initial detection



object model (histogram)

$$H_{model} = \{q_u\}_{u=1...n} \qquad \sum_{u=1}^{n} q_u = 1$$

histogram backprojection

input image, I(x)

histogram backprojection

backprojection image

$$w(x) = \sqrt{\frac{H_{model}(I(x))}{H_{bkg}(I(x))}}$$

mean-shift localization

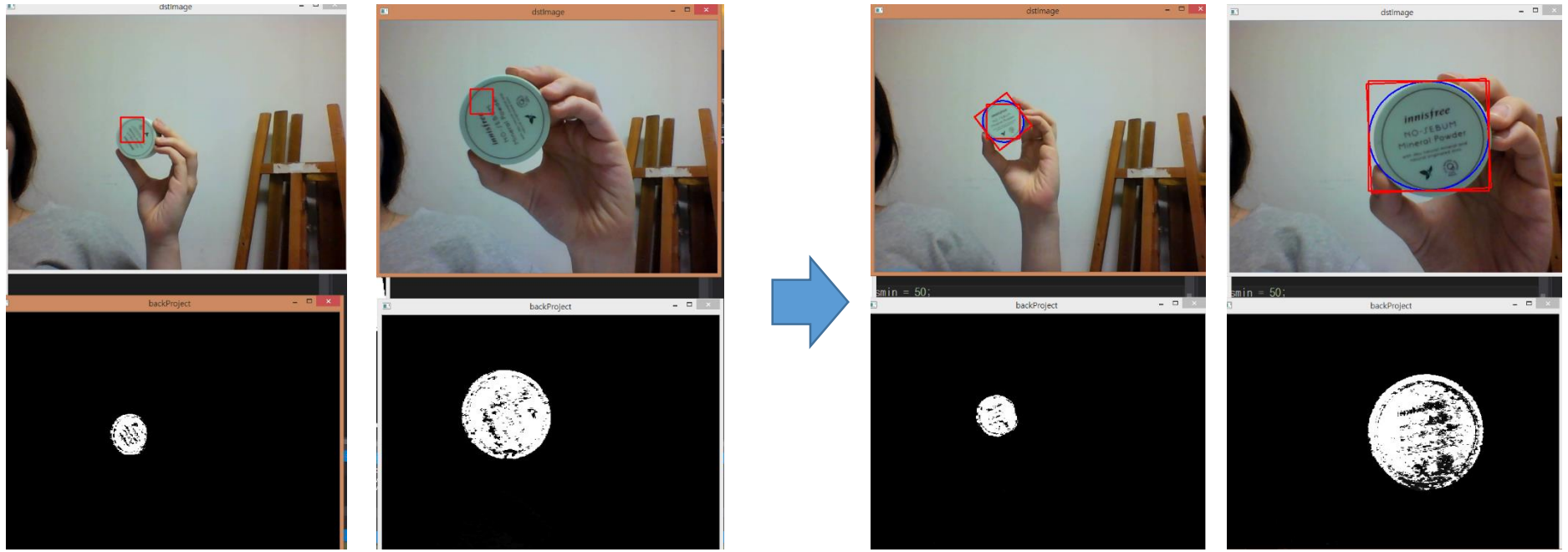
$$\Delta x = \frac{\sum w(x)(x - \bar{x}_{old})K(x - \bar{x}_{old})}{\sum w(x)K(x - \bar{x}_{old})}$$

$$\bar{x}_{new} = \bar{x}_{old} + \Delta x$$

mean-shift localization (density mean)

Camshift

- Cam shift
  - Modified version of mean-shift
  - The size of search window can be changed

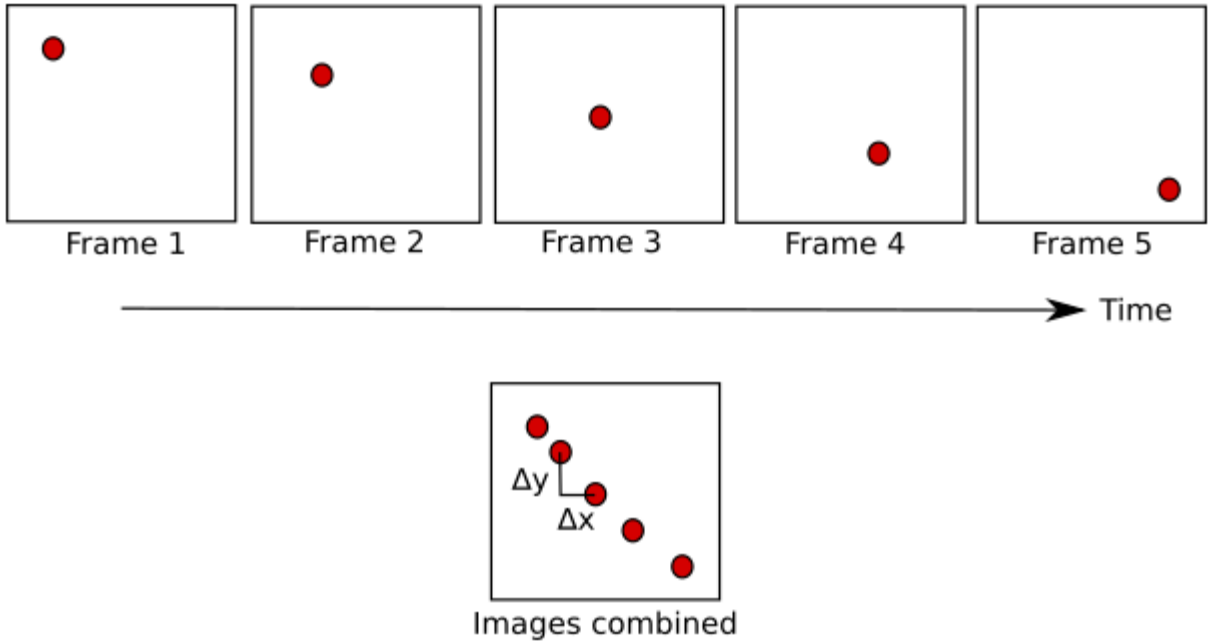


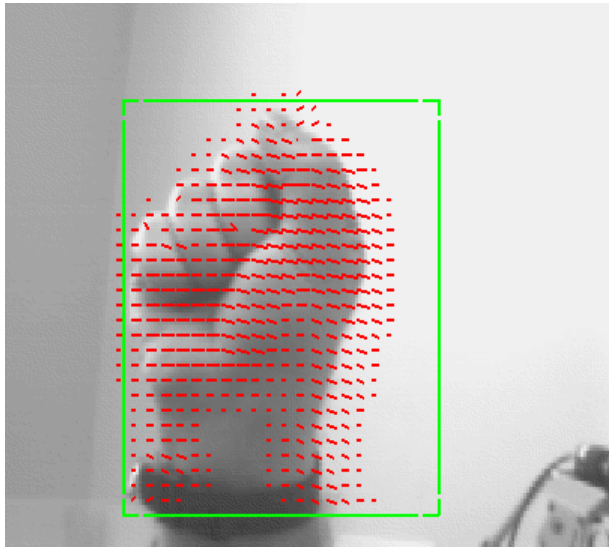
Mean-shift

Cam shift

Optical Flow

- Optical flow is the apparent motion of brightness patterns in the image





Optical Flow

- KLT algorithm
  - Assumption
    - Intensity of objects are not changed over consecutive frames
    - Movement of pixels are similar to that of adjacent pixels

➔ $I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t)$

- By applying Taylor series
  - $I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x, y, t) + \frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t$

➔$\frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t = 0$

- Extract features first and track the extracted features

Optical Flow

- KLT algorithm with pyramids
  - Original KLT algorithm cannot handle large movement
  - To overcome this limitation, image pyramid is used