Salient Object detection in a video using local Steering Kernel

by

Yogesh Kumar

M.Tech (SPDD)

Delhi Technological University

Outline

- Introduction/Motivation
- Local steering kernel
- System overview
- Shot detection
- Result
- Limitation/ Possible solutions.

Motivation

- Information abundance
- Growth of digital video data
- Lack of tools for classify and
 retrieve video content

Content based image retrieval

- Search for digital images in large databases
- Analysis the actual contents of the image
- Image retrieval feature extraction or content based algorithm.
- Content based algorithm
  - colors,
  - Shapes
  - Textures etc.

Content-based Video indexing

- The process of attaching content based labels to video shots
- Essential for content-based classification and retrieval
- Using automatic analysis techniques

  - shot detection, video segmentation

  - key frame selection

  - object segmentation and recognition

  - visual/audio feature extraction

  - speech recognition, video text, VOCR

- Content-based Video Classification
  - Segment & classify videos into meaning categories
  - Classify videos based on predefined topic
  - Useful for browsing and searching by topic
  - Multimodal method
    - Visual features
    - Audio features
    - Motion features
    - Textual features
  - Domain-specific knowledge

Content-based Video Retrieval

- Simple visual feature query
  - Retrieve video with key-frame: Color-R(80%),G(10%),B(10%)
- Feature combination query
  - Retrieve video with high motion upward(70%), Blue(30%)
- Query by example (QBE)
  - Retrieve video which is similar to example
- Localized feature query
  - Retrieve video with a running car toward right
- Object relationship query
  - Retrieve video with a girl watching the sun set
- Concept query (query by keyword)
  - Retrieve explosion, White Christmas

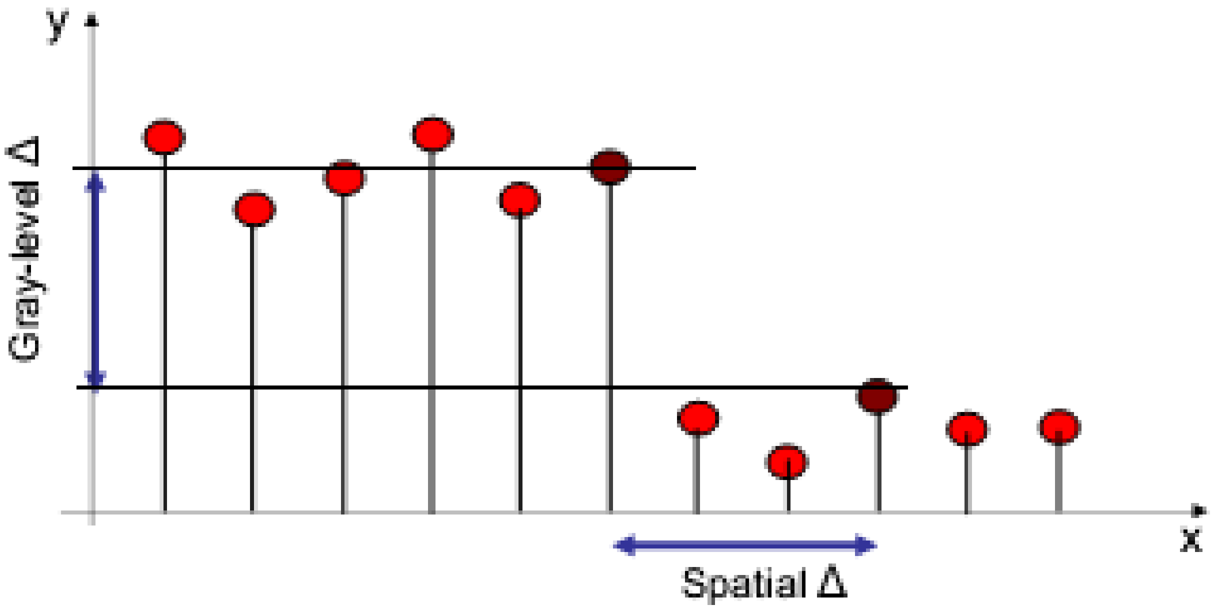
- Feature Extraction
  - Color features
  - Texture features
  - Shape features
  - Sketch features
  - Audio features
  - Camera motion features
  - Object motion features

Objective

- New feature extraction
- Training free detection

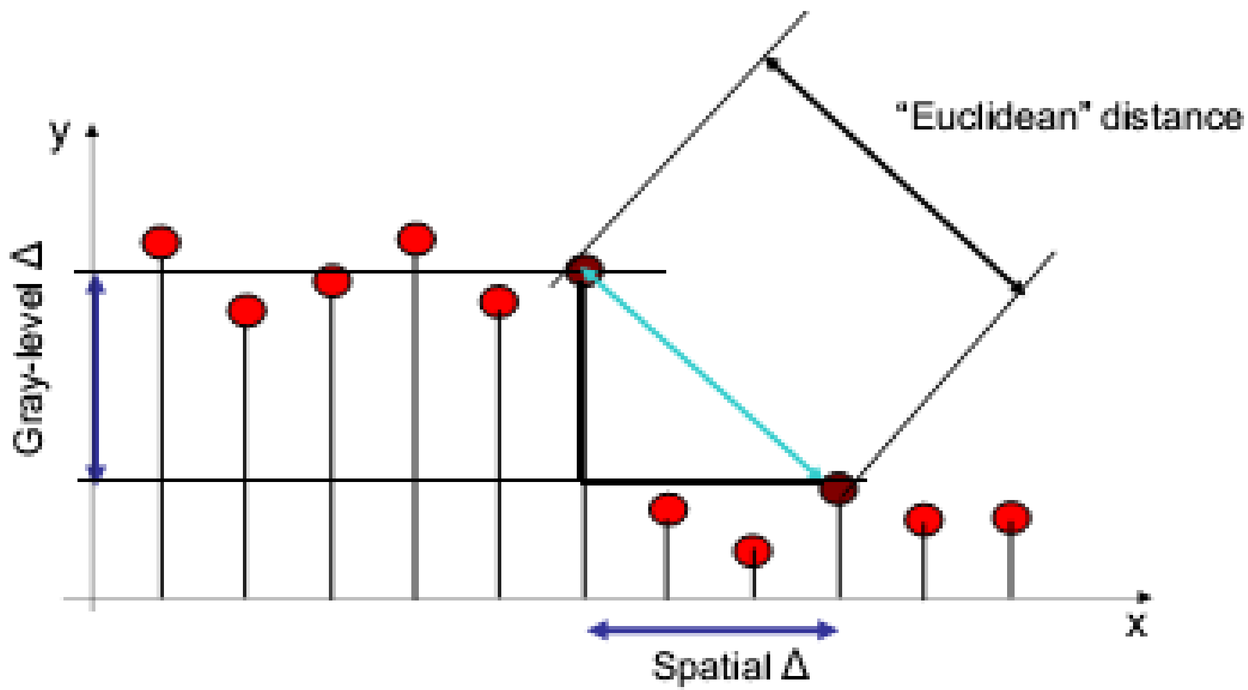
Defining a point-wise measure

- To measure the similarity of two pixels, consider
  - Spatial distance
  - Gray-level distance



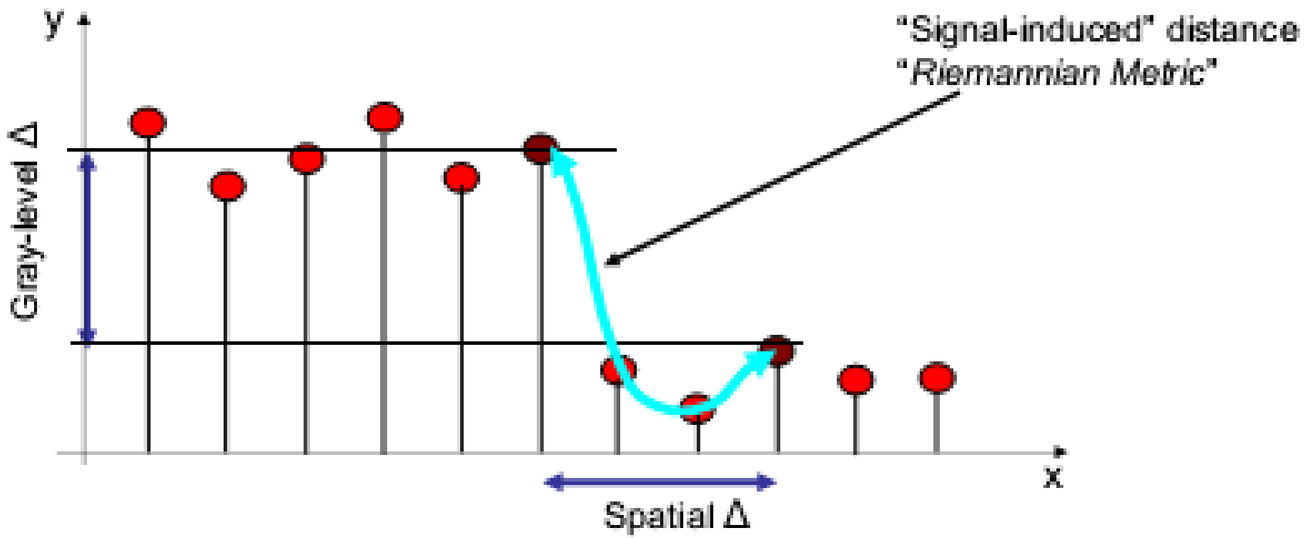
Simplest measures

- Basic ways to incorporate the two Δs:
  - Bilateral kernel
  - Non-local Means kernel



Defining a measure

- More effective ways to combine the two Δs:
  - LARK kernel [Takeda, et al. '07]
  - Beltrami kernel [Sochen, et al. '98]



A digression to Non-parametric Kernel Regression

- The data fitting problem

Zero mean, i.i.d. noise

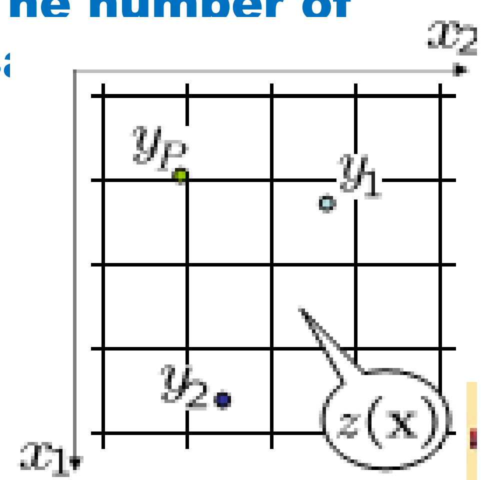
$$y_i = z(x_i) + \varepsilon_i, \quad i = 1,2,\ldots,P$$

Given Samples

The Regression Function

The Sampling Position

The number of s



The particular form of z(x) may remain unspecified for now.

Locality in Kernel Regression

- The data model

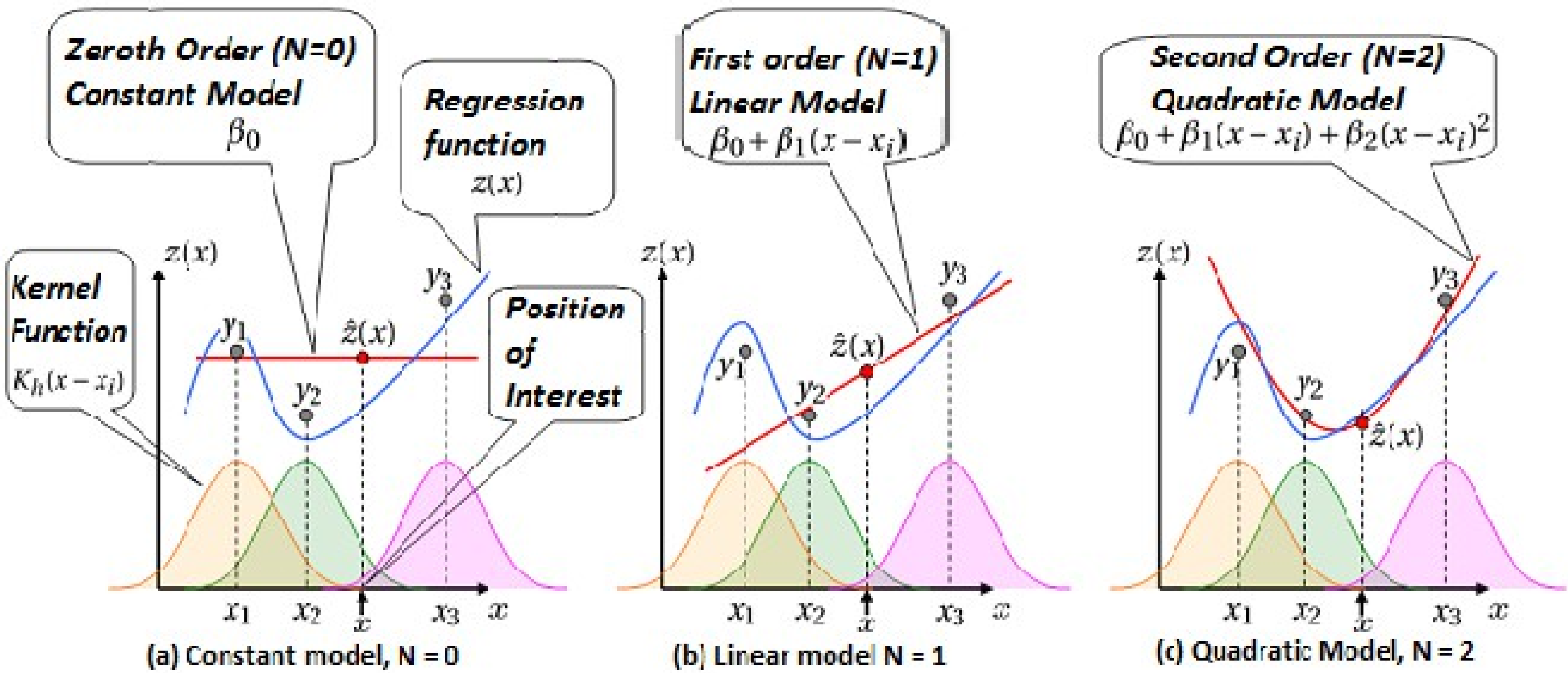
$$y_i = z(x_i) + \varepsilon_i \,,\; i = 1,2,\ldots,P$$

- Local representation (N-term Taylor expansion)

$$z(x_i) \approx z(x) + \{\nabla z(x)\}^T (x_i - x) + \frac{1}{2}(x_i - x)^T \{H z(x)\}(x_i - x) + \ldots$$

$$z(x_i) \approx \beta_0 + \beta_1^T (x_i - x) + \beta_2^T \, vech\{(x_i - x)(x_i - x)^T\} + \cdots$$

Unknowns


Zeroth Order (N=0)
Constant Model
β0
Regression function
z(x)
Kernel Function
Kh(x − xi)
Position of Interest
First order (N=1)
Linear Model
β0 + β1(x − xi)
Second Order (N=2)
Quadratic Model
β0 + β1(x − xi) + β2(x − xi)²
(a) Constant model, N = 0
(b) Linear model N = 1
(c) Quadratic Model, N = 2

- We have a local representation with respect to each sample:

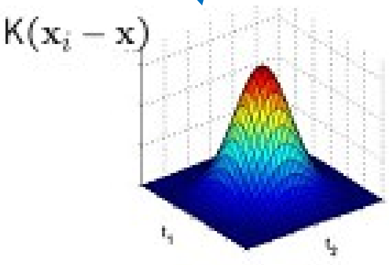
$$y_1 = \beta_0 + \beta_1^T (x_1 - x) + \beta_2^T vech\{(x_1 - x)(x_1 - x)^T\} + \cdots + \epsilon_1$$
$$y_2 = \beta_0 + \beta_1^T (x_2 - x) + \beta_2^T vech\{(x_2 - x)(x_2 - x)^T\} + \cdots + \epsilon_2$$
$$\vdots \qquad\qquad \vdots$$
$$y_p = \beta_0 + \beta_1^T (x_P - x) + \beta_2^T vech\{(x_P - x)(x_P - x)^T\} + \cdots + \epsilon_P$$

- Optimization

$$\min_{\{\beta_n\}_{n=0}^{N}} \Sigma_{i=1}^{P} \left[y_i - \beta_0 - \beta_1^T (x_i - x) - \beta_2^T vech\{(x_i - x)(x_i - x)^T\} - \ldots \right]^2 K_H(x_i - x)$$

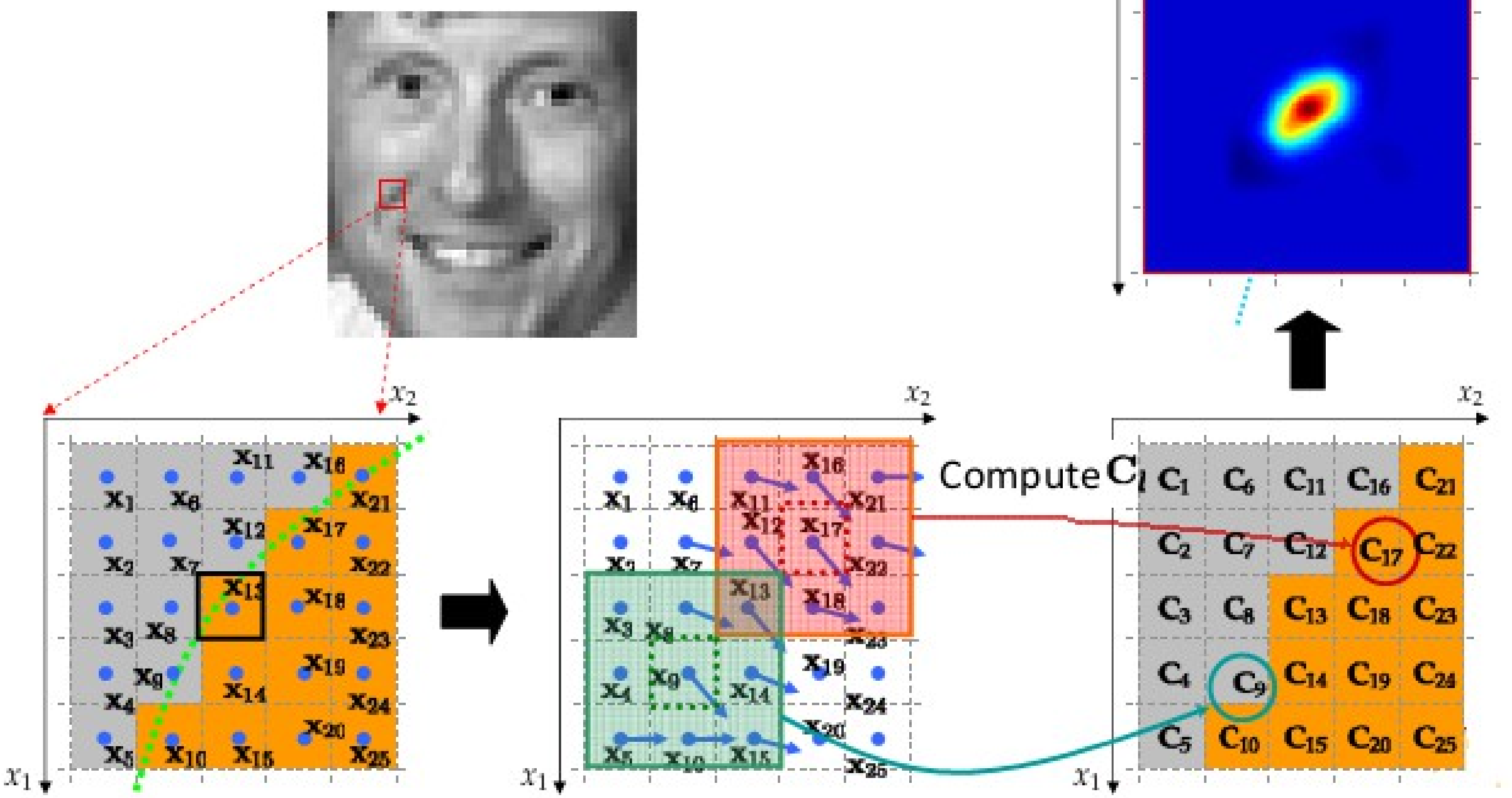
This term give the estimated pixel value z(x)

Choice of kernel function is open, e.g. Gaussian

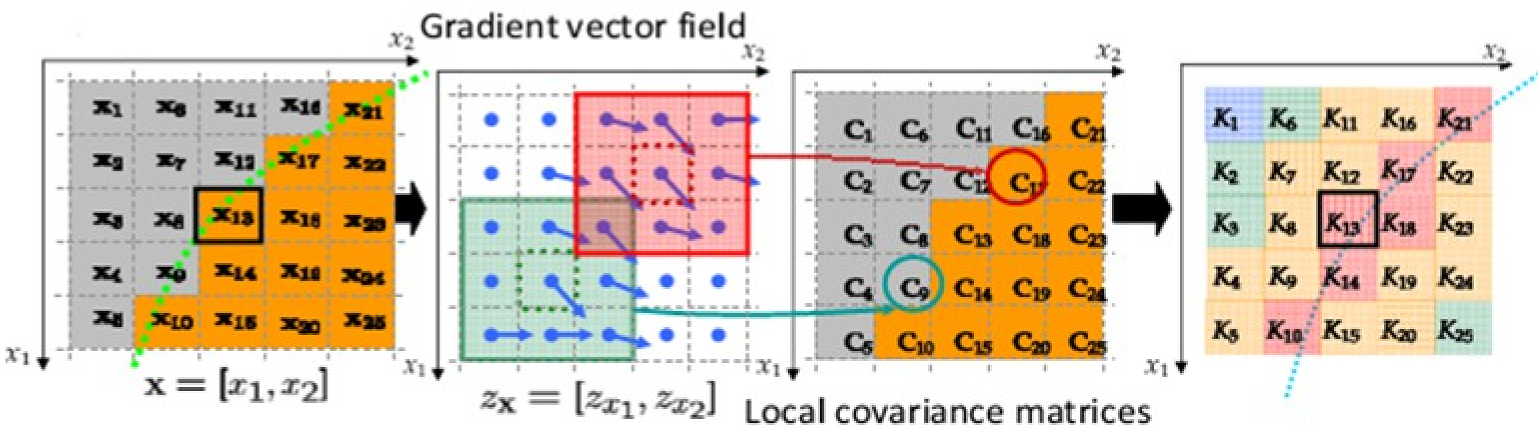


LARK Kernels

$$K(\mathbf{C}_l, \mathbf{x}_l, \mathbf{x}) = \exp\left\{-(\mathbf{x}_l - \mathbf{x})'\mathbf{C}_l(\mathbf{x}_l - \mathbf{x})\right\}$$



- Gradient matrix over a local patch

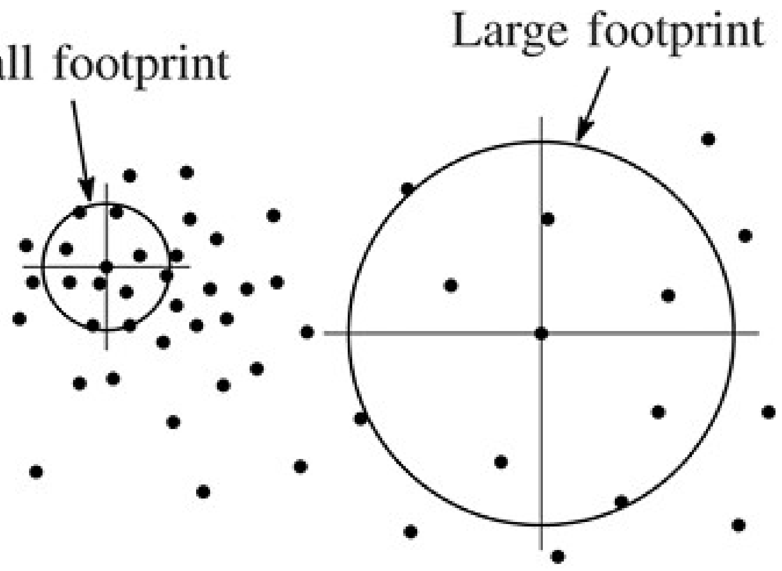


Locally Adaptive Regression Kernel: LARK

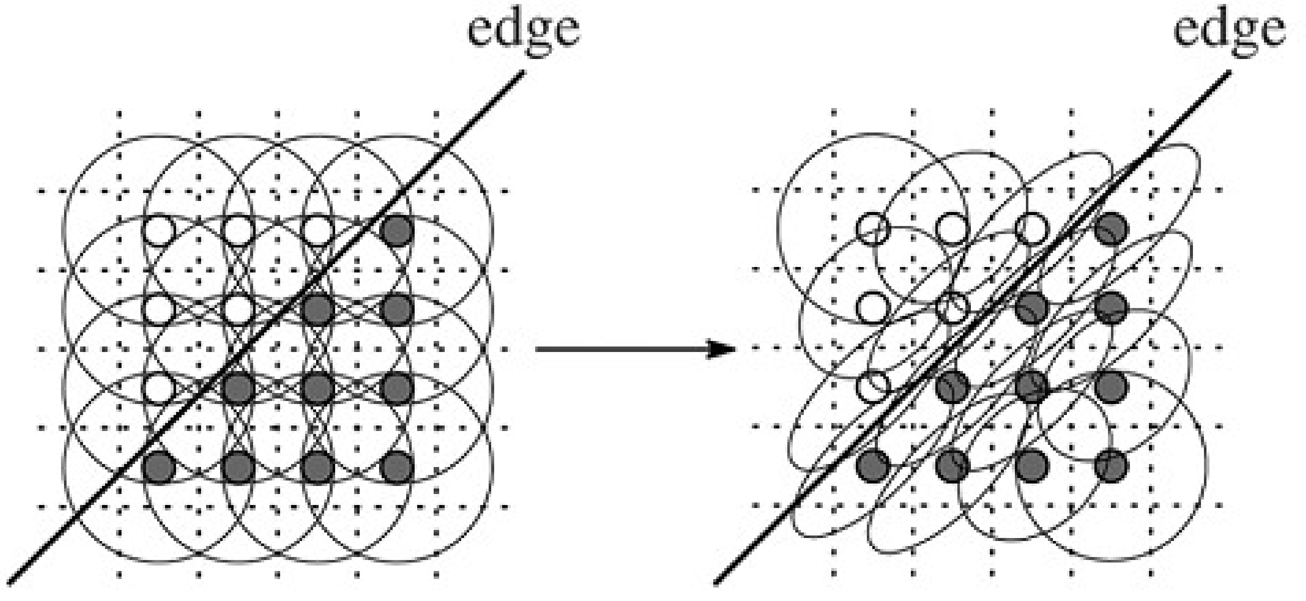
$$K(\mathbf{C}_l, \mathbf{x}_l, \mathbf{x}) = \exp\left\{-(\mathbf{x}_l - \mathbf{x})'\mathbf{C}_l(\mathbf{x}_l - \mathbf{x})\right\}$$

$$\hat{c}_i \approx \begin{bmatrix} \sum_{x_j \in w_i} f_x(x_j) f_x(x_j) & \sum_{x_j \in w_i} f_x(x_j) f_y(x_j) \\ \sum_{x_j \in w_i} f_y(x_j) f_x(x_j) & \sum_{x_j \in w_i} f_y(x_j) f_y(x_j) \end{bmatrix}$$

$$H_i^{steer} = h\,\mu_i\,C_i^{\frac{1}{2}}$$

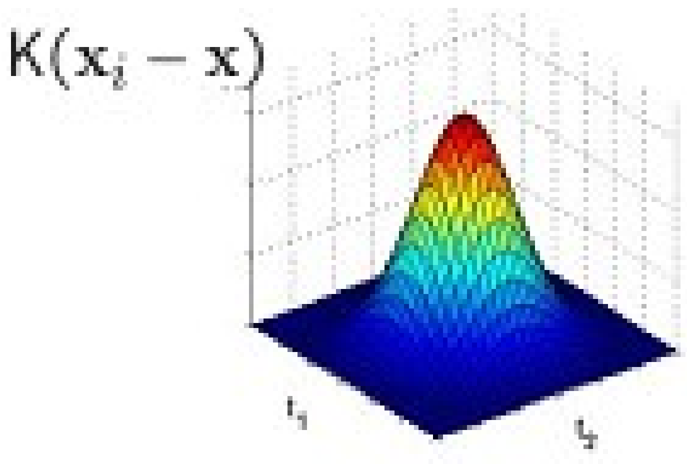


- Kernel Comparison



- Choice of kernel function

$$K_{H_i^{steer}}(\boldsymbol{x_i}-\boldsymbol{x})=\frac{\sqrt{det(C_i)}}{2\pi h^2 \mu_i^2}\exp\left\{\frac{-(x_i-x)^T C_i (x_i-x)}{2h^2\mu_i^2}\right\}$$



Object Detection

- LSK obtained are given by
  - For query- $K_Q(x_i - x : H_i)$
  - For target- $K_T^j(x_i - x : H_i)$

- Data normalization and PCA

$$W_Q(x_i - x) = \frac{K_Q(x_i - x ; H_i)}{\sum_{l=1}^{P^2} K_Q(x_l - x ; H_l)} \longrightarrow F_Q$$

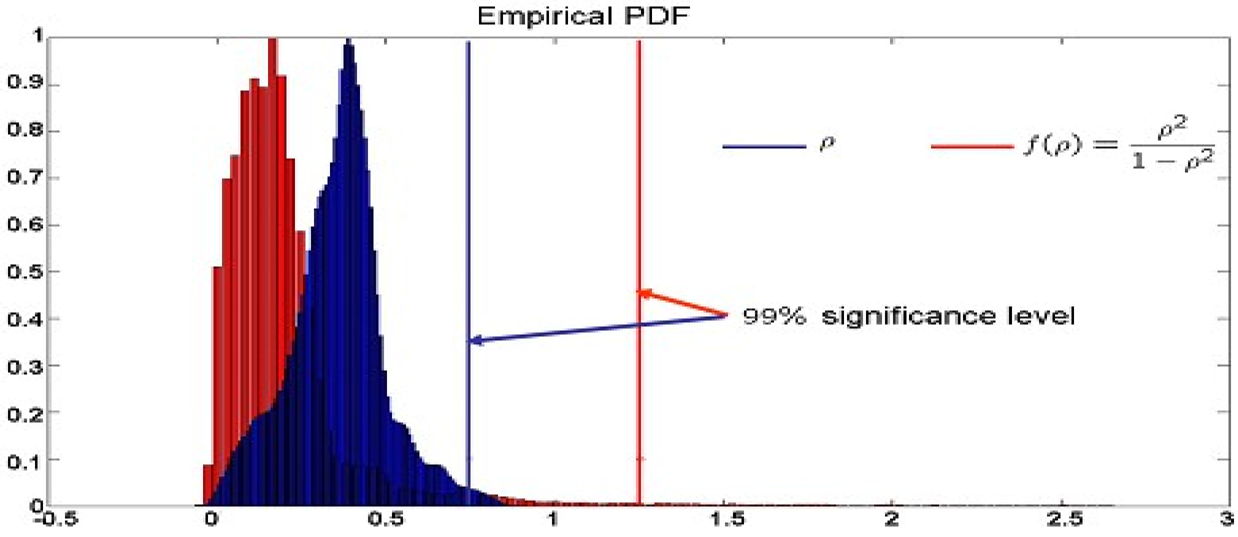
$$W_T^j(x_i - x) = \frac{K_T^j(x_i - x ; H_i)}{\sum_{l=1}^{P^2} K_T^j(x_l - x ; H_l)} \longrightarrow F_T$$

- Similarity of patch in target to the query

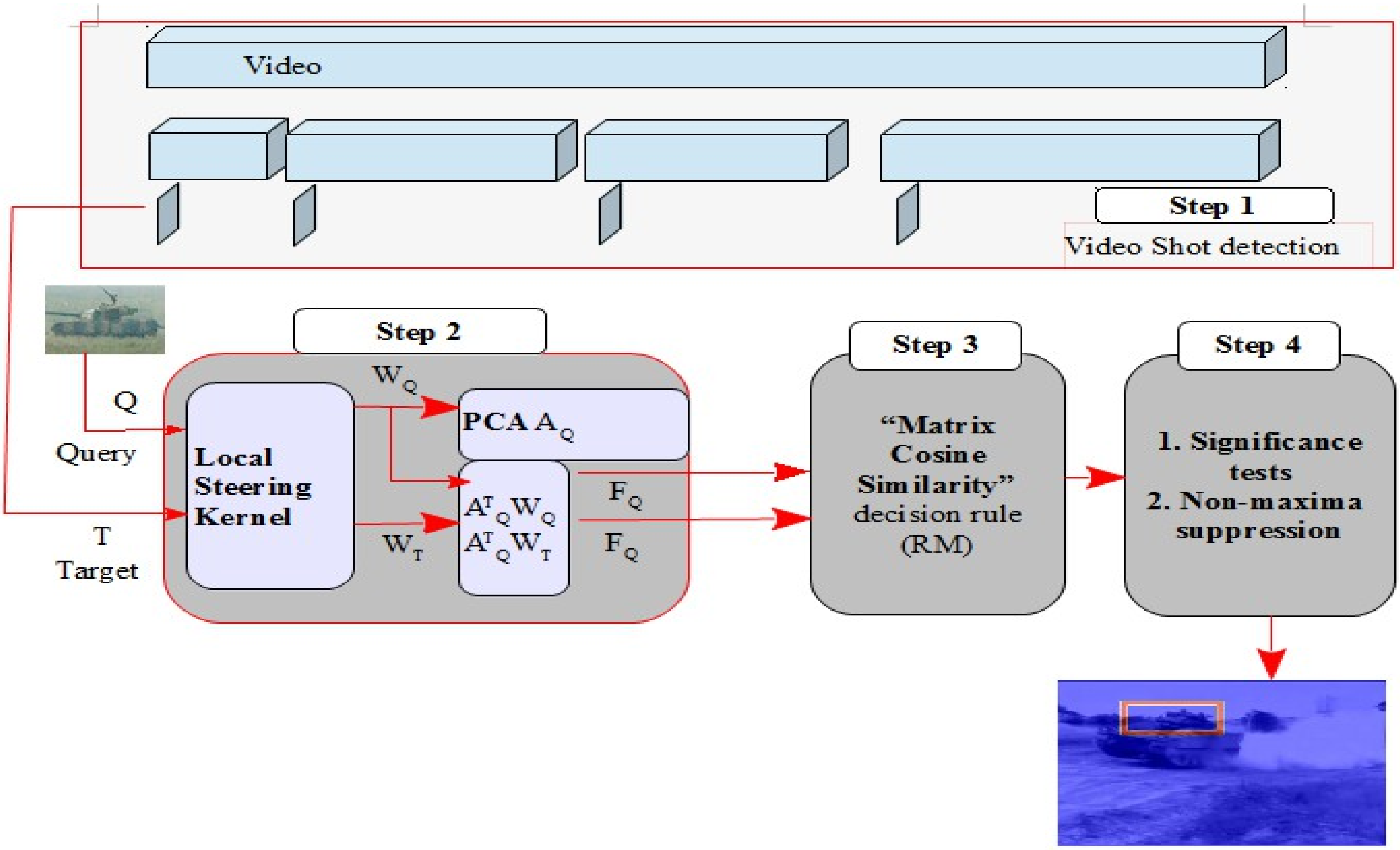
$$\rho_j = \left\langle \frac{F_Q}{\|F_Q\|}, \frac{F_Q}{\|F_Q\|} \right\rangle$$

- Resemblance map

$$f(\rho_j) = \frac{\rho_j^2}{1-\rho_j^2}$$



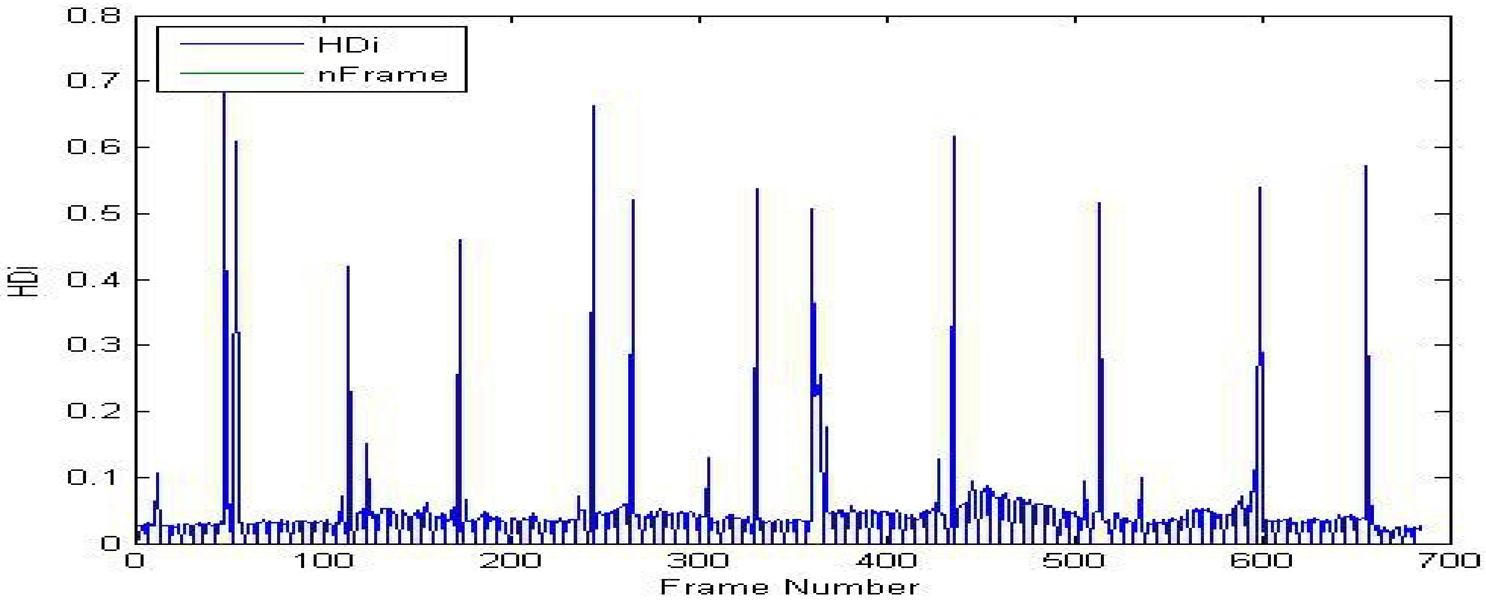
Methodology



Histogram intersection

- Assumption- Abrupt change in frames

$$HDi = 1 - (1/3\mathrm{n}) \times [\Sigma_{j=1}^{n} min(F_{rj}^{i}, F_{rj}^{i+1}) + \Sigma_{j=1}^{n} min(F_{gj}^{i}, F_{gj}^{i+1}) + \Sigma_{j=1}^{n} min(F_{bj}^{i}, F_{bj}^{i+1})]$$



- Query image

Pixel size 100 x 54

Pixel size 60 x 41

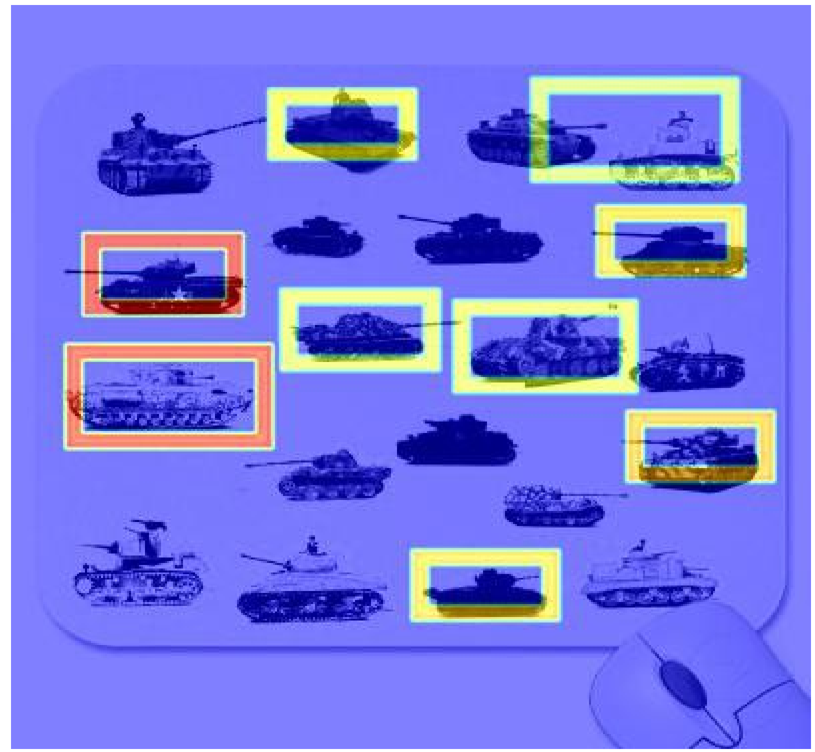
- Target image
  - Standard size of video image of 320x240 or 640x320..

Result

Limitations

- Object size
  - Size of object in the target image could be large
- Orientation
  - Different orientation of objects
- Time taken

Chained Query Image example

Conclusion and future work

- Powerful training-free non-parametric object detection framework
- Capture underlying data structure
- Automatically detect in the target image the presence, the number, as well as location of similar objects to the given query image
- Proposed framework operates using a single example of an image of interest to find similar matches
- Does not require any segmentation or pre-processing step of the target image