CS365 Project

Person Detection and Tracking

Vimal Sharma
Shikhar Sharma

Supervised by -
Dr. Amitabha Mukerjee

Motivation/Intent

- Pedestrian accidents are an issue of serious concern throughout the world.
- A rich framework built into the vehicle to detect and track pedestrians on road can help avoid fatalities.

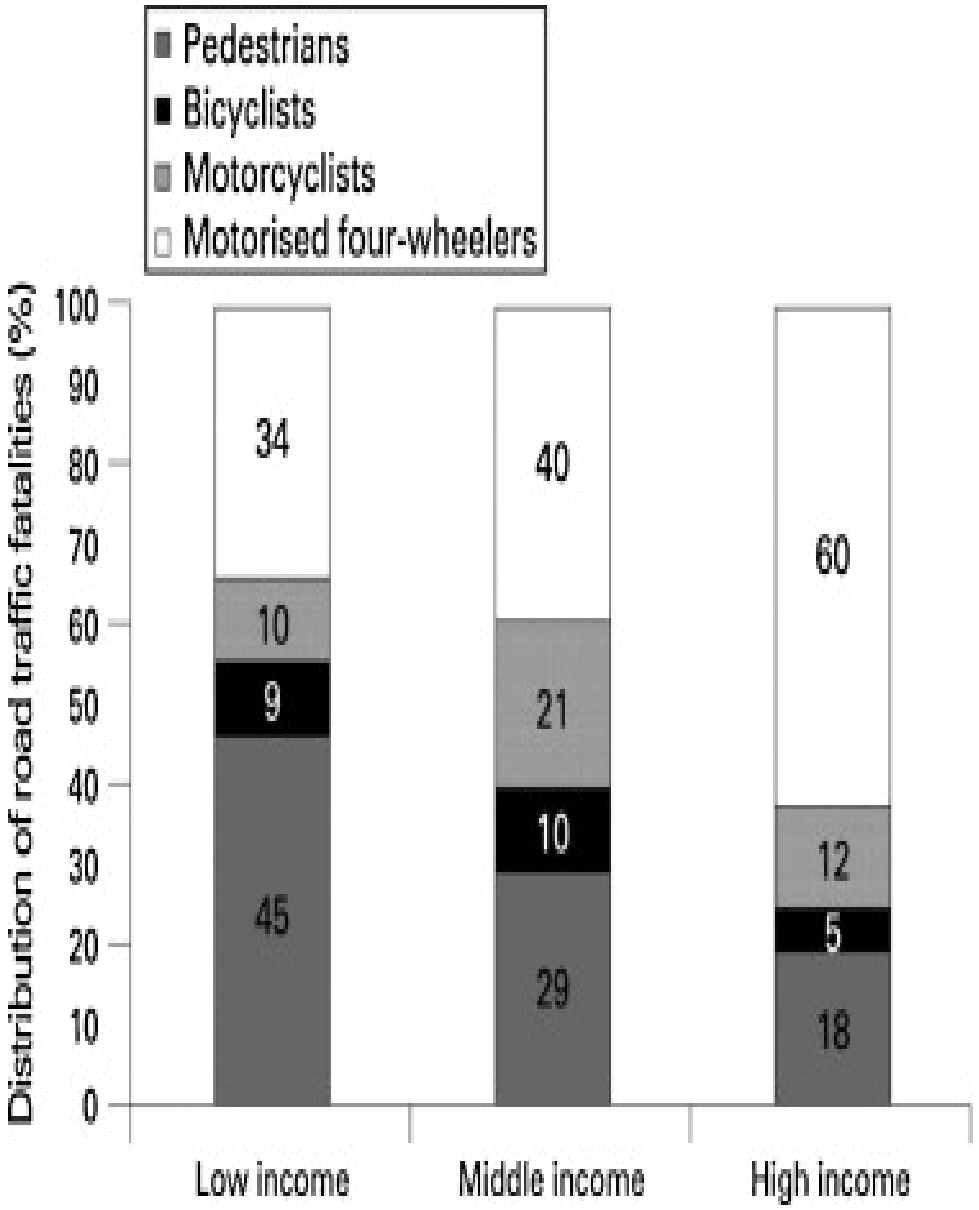


Image Source – Naci etal. 2009

Motivation/Intent

- We present a framework for person detection from a video stream.
- On successful detection, we track the object throughout the video.

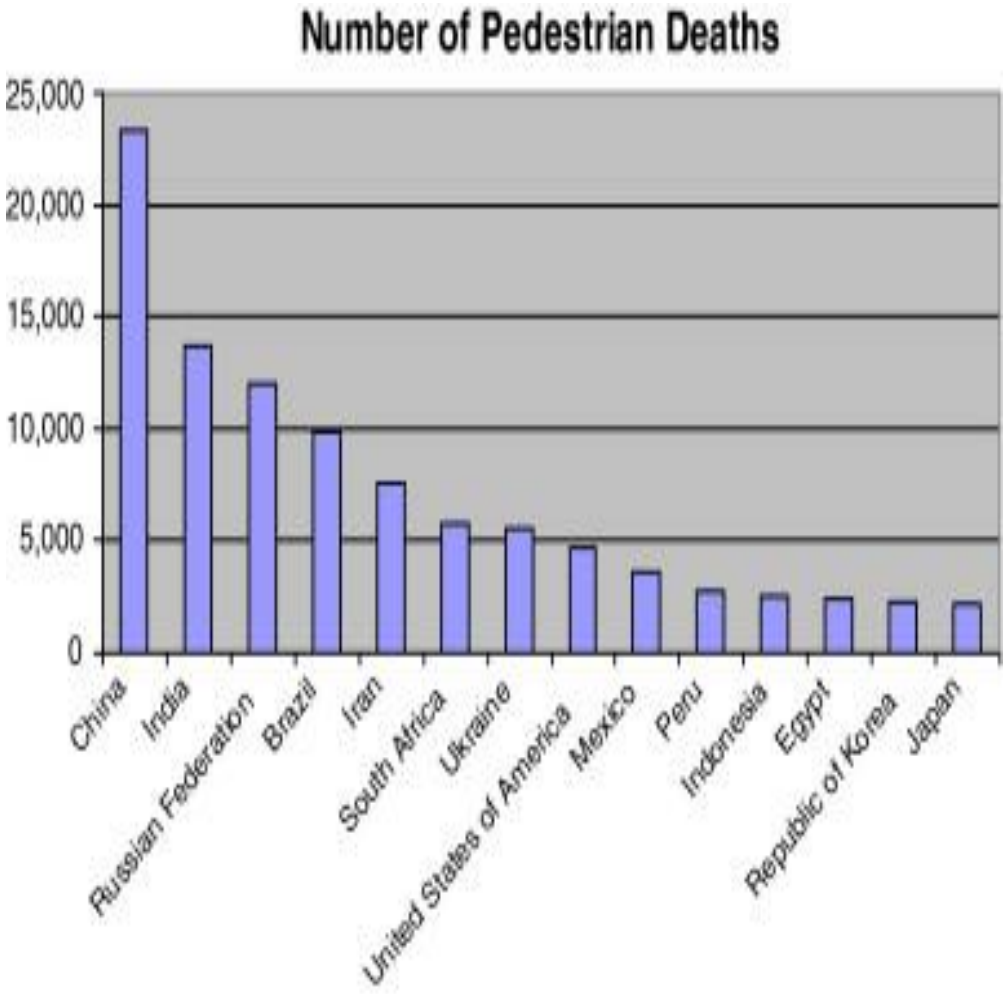
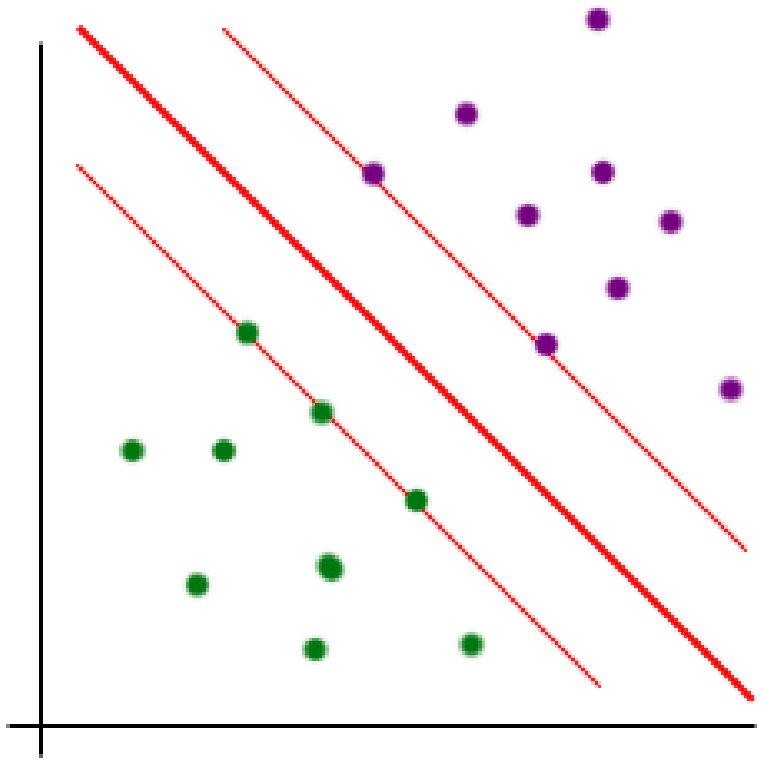


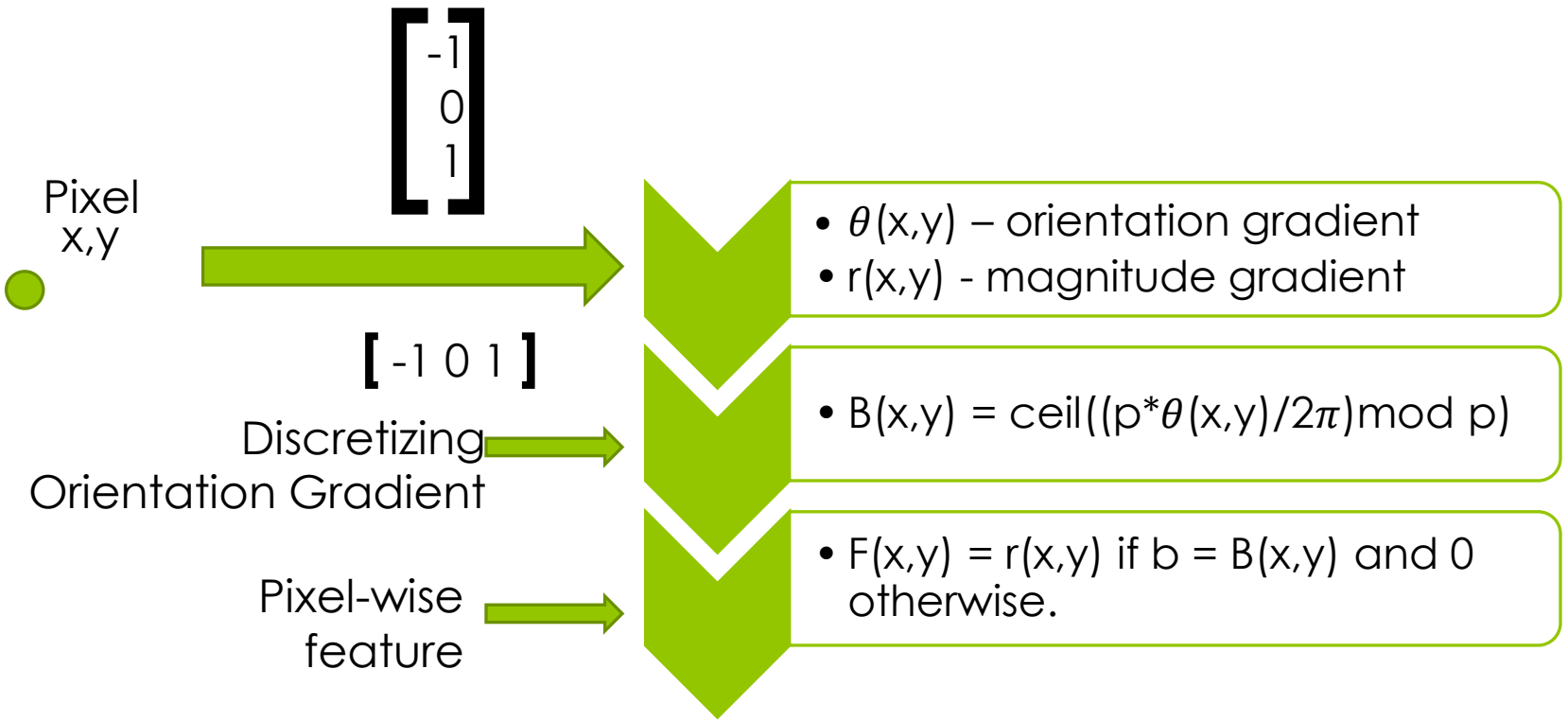
Image Source - WHO 2009

Methods Used

- Histogram of Oriented Gradients(HOG)
- Principal Component Analysis(PCA)
- Part based Deformable Models(Pictorial Structures) for Object Recognition
- Latent SVM
- Background Subtraction using Approximate Median



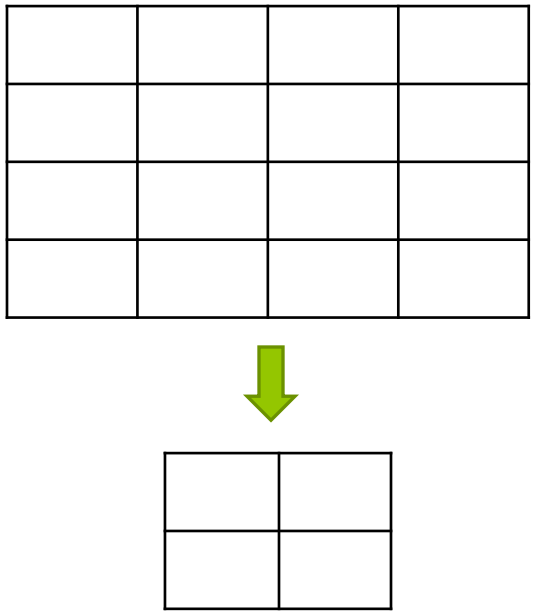
HOG Features



p is the number of bins
F(x,y) is the p-dimensional pixel wise feature

Spatial Aggregation of Features

To reduce the size of the computed feature map, the pixel level feature maps are aggregated to “cell” level feature maps.



Normalization

We obtain a 36 dimensional HOG feature after subjecting C[i,j] to four normalizations.

36 dimensional HOG feature → PCA → 11 dimensional feature vector

Latent SVM

- It is called latent because the training images are not completely annotated – the part locations for different body parts are not given. Only the bounding box for the complete person is given.

- The classifier scores the examples using the function

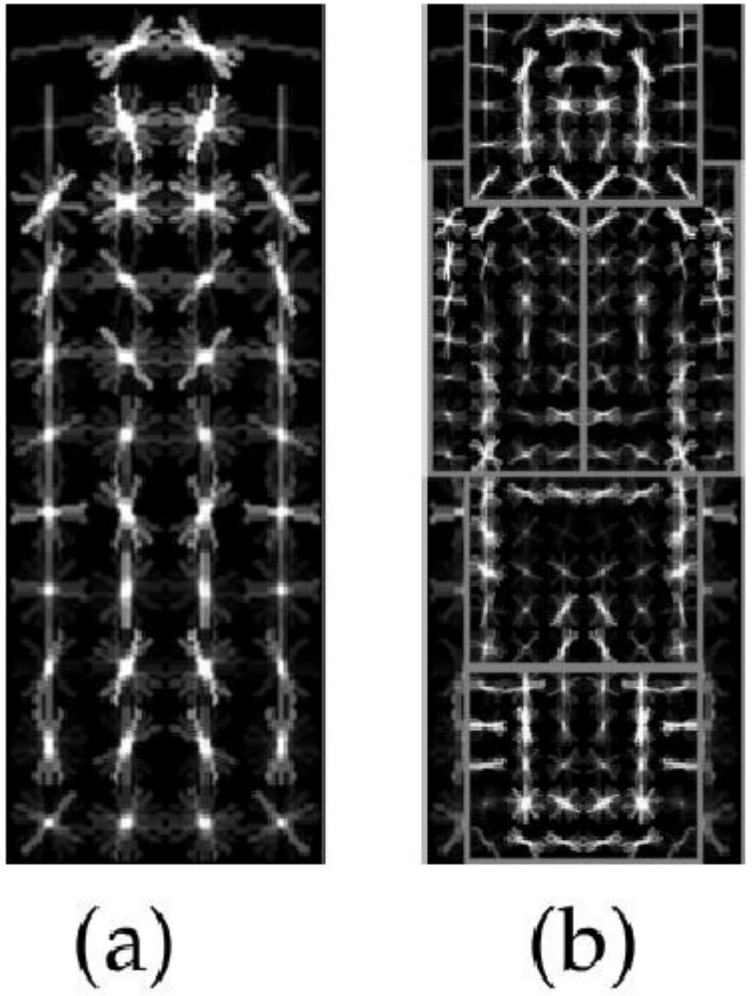
$$f_w(x) = \max_{z \in Z(x)} w.\Phi(x,z)$$

w - vector of model parameters
z - latent values.
Z(x) - defines the possible latent values for an example x.
Φ(x,z) – feature vector

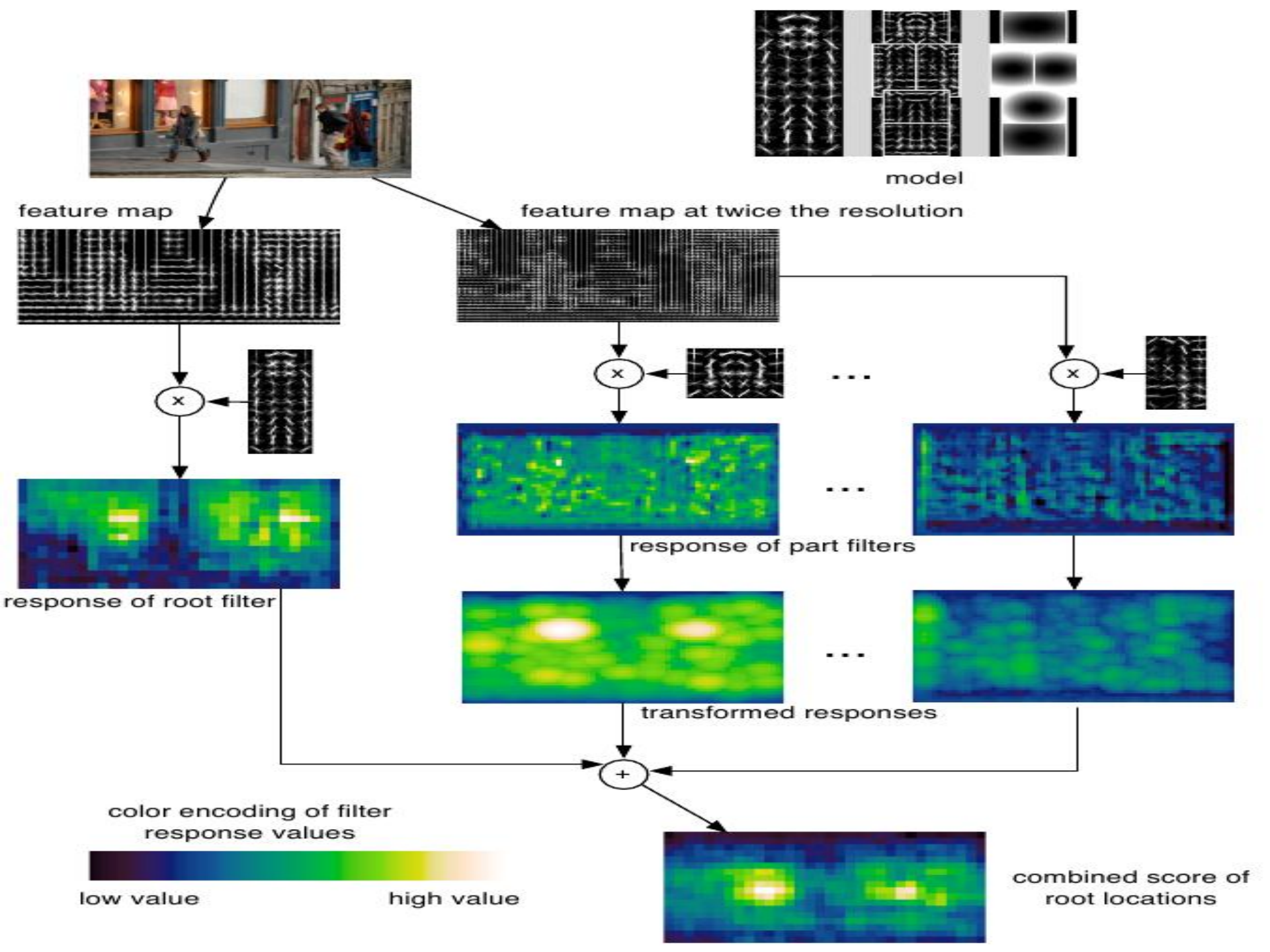
Human Body Model

The resulting model consists of

(a) A root filter – coarsely describes the configuration of the entire object
(b) A set of part based filters - describe individual parts of the object more precisely.



(a) (b)


model
feature map
feature map at twice the resolution
response of part filters
response of root filter
transformed responses
color encoding of filter
response values
low value
high value
combined score of
root locations

Bounding Box

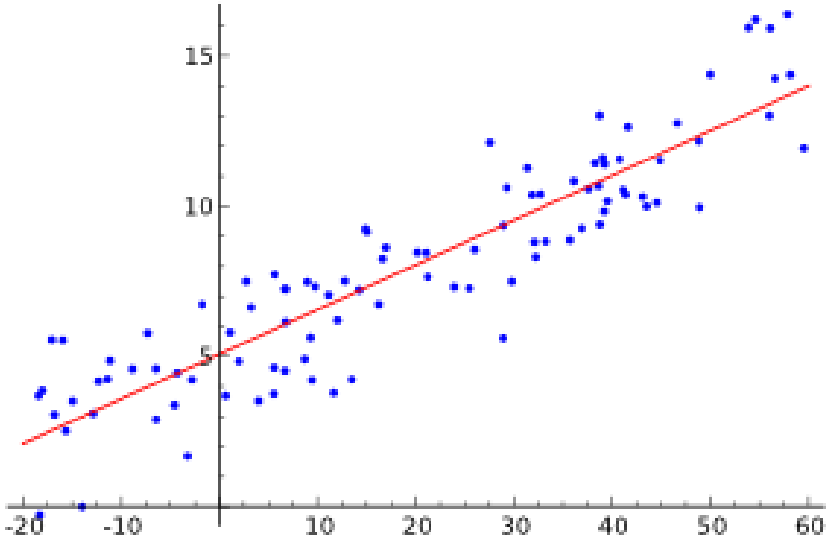


Image Source : Wikipedia

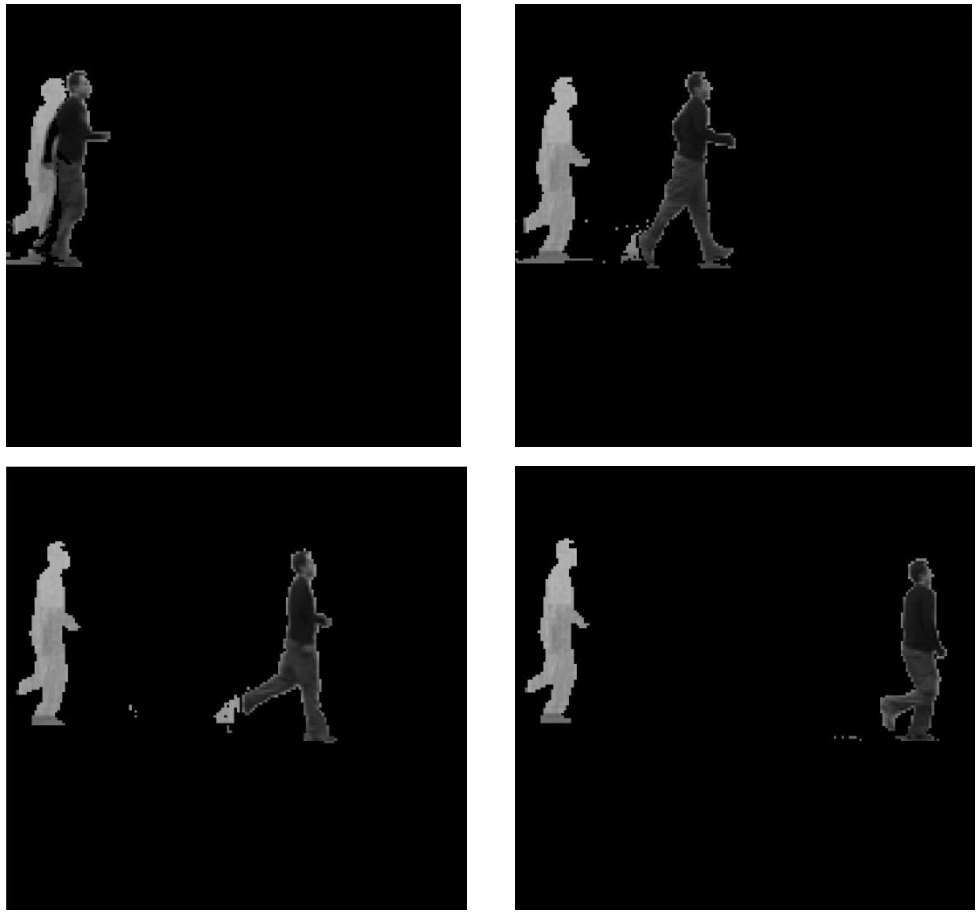
Bounding Box Prediction

By linear least squares regression the upper left and the lower right corner positions of the bounding box are predicted.

TRACKING

On a successful detection of person in video stream the tracker tracks the person throughout the video stream. We use background subtraction using Approximate Median method.

Stills of output on the Weizmann video dataset



Code & Datasets Used

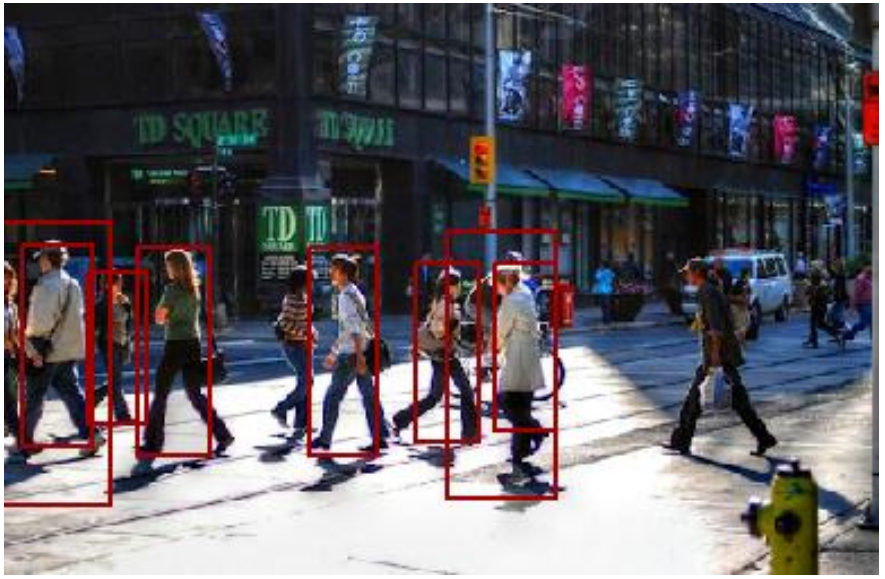
- Code for detection obtained from project page of [1]
- PASCAL VOC 2006, VOC 2009 Datasets
- Weizmann video Dataset
- Implementation of background subtraction by us in Matlab

The success rate for human detection on the PASCAL VOC 2009 dataset is 43.8%.

INRIA Person average precision is 88.2 .

Persons were detected in the 1st frame in the Weizmann Running and Walking Dataset

Detection Results



References

[1] **Pedro F. Felzenszwalb, Ross B. Girshick, David McAllester and Deva Ramanan[2010]** Object Detection with Discriminatively Trained Part-Based Models. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 32, NO. 9, SEPTEMBER 2010

[2] **P.F. Felzenszwalb and D. P. Huttenlocher[2005]** Pictorial structures for object recognition. IJCV 2005.