# Human Activity Recognition

Satwik Kottur [1], Dr. Subhasis Chaudhuri [1]
Indian Institute of Technology Bombay

## Abstract

With the increasing rise in the need for security and surveillance, particularly in crowded areas like airports, shopping malls and social gatherings, the problem of human detection and activity recognition has attained importance in the vision community. This is both an interesting and challenging problem due to two reasons:

• The behavior can be governed by set of underlying rules which when exploited give insights into the crowd motion. For instance, motion analysis of players in a game of basketball strictly follows the rules of the game.

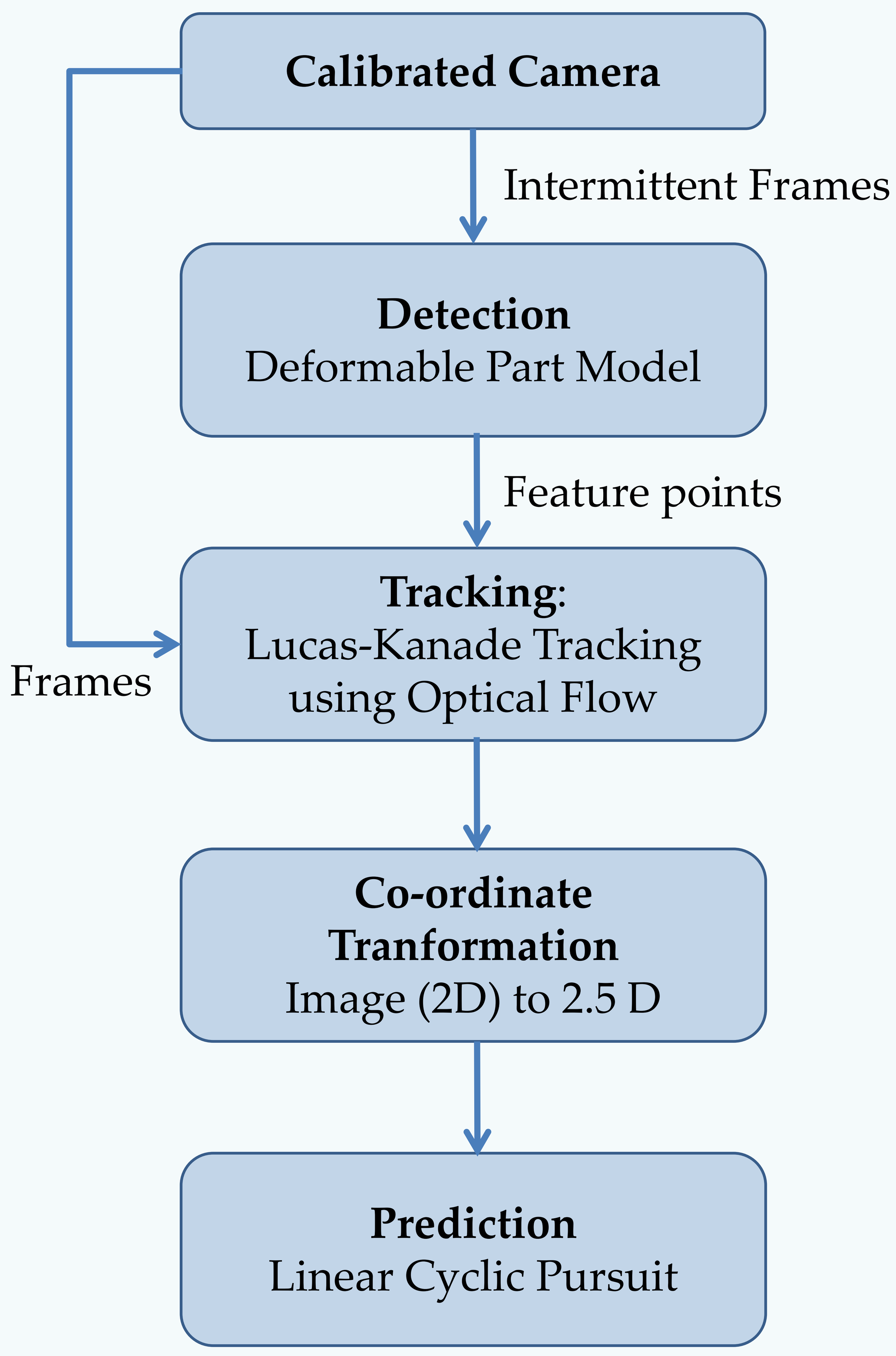
• Different levels of analysis give different inferences. Motion of people in a busy street, when seen as individual agents, might look completely random and uncorrelated. However, a crowd level analysis might reveal interesting patterns like crossing the road in opposite directions, halting at the traffic signs, etc.

The problem of human activity has been studied primarily using holistic and reductionist approaches. In the former approach, crowd is modeled as a single entity and behavior is analyzed. The latter approach considers sparse individual motions and tracks humans by applying suitable filters to evaluate motion parameters.

Scope :
Study of various existing techniques by getting them together into a working pipeline for human detection and analysis in a social gathering.

## Approach

A single, calibrated camera is used to obtain the video from a social gathering, which is the input to the system. Starting from this video, we detect humans, track them using feature point and predict the short term trajectory.




Each module is explained in following sections.

## Detection

Considers articulated nature of human body into account for detection.

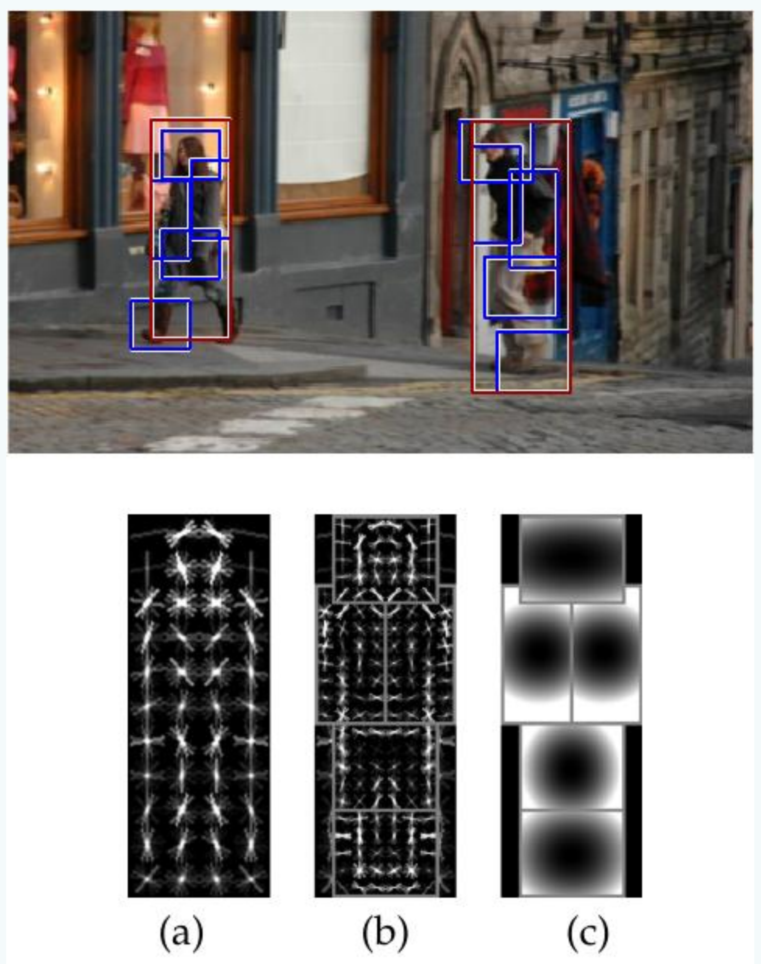



Fig: Detection using Deformable part model (a) Root filter (b) Part filter (c) Location model *Courtesy: [1]*

•Learns linear models (SVMs) for parts, root and part locations with respect to the root, using supervision
•Histogram of Gradients (HOG) features used
•Scores root detection, part detections and part displacements from expected locations
•Threshold on score followed by Non-Maximal Suppression (NMS) to reduce false positives
•Location of parts and root used as feature points in tracking

## Tracking

•Detection is computational expensive
•Hence, done for intermittent frames
•To satisfy the real-time execution constraint, tracking using Lucas-Kanade algorithm is used
•Tracking of feature point is based on the optical flow of the image

## Co-ordinate Transformation

•For any meaningful inferences to be made, detected 2D co-ordinates from the image plane must be transformed to world co-ordinates
•The concept of 2.5D as used by Neha et al. [2] with assumption of an average height plane
•Since the camera is calibrated, we can back-project the point to find the appropriate 2.5D co-ordinate of P1, as shown.

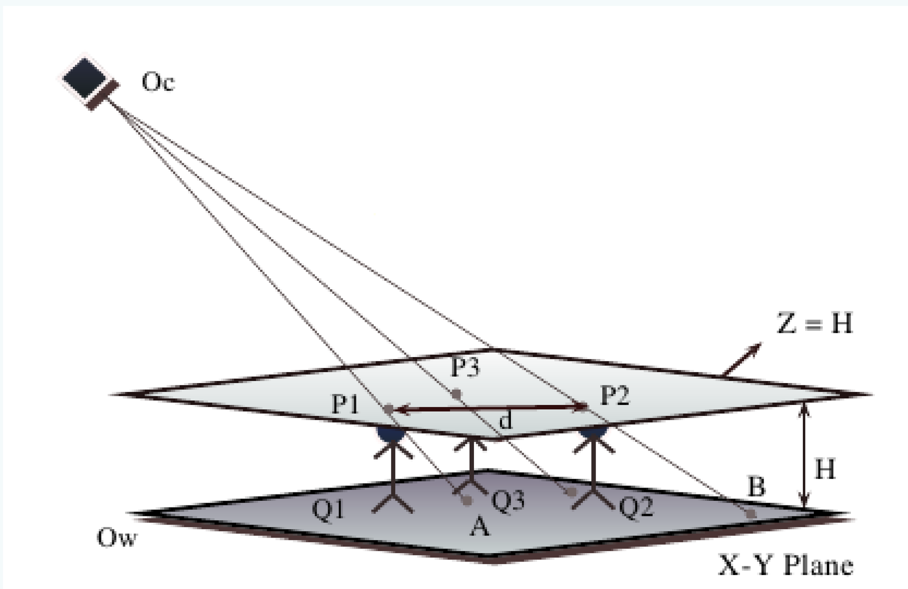



Fig: Illustration of co-ordinate transformation *Courtesy: [2]*

$$x_{p1} = \frac{H - z_c}{z_a - z_c}(x_a - x_c) + x_c$$

$$y_{p1} = \frac{H - z_c}{z_a - z_c}(y_a - y_c) + y_c$$

$$z_{p1} = H$$

## Prediction

After evaluating the world co-ordinates, linear cycle pursuit model is used to predict short-term trajectories.

For a group of N agents and their motion in $\mathbf{R}^2$, model works on the following principles:

1. Motion in both dimensions are independent of each other i.e. X direction motion does not have a bearing on Y direction motion.
2. Direction of motion for an agent is towards the weighted mean of positions of other agents.

$$\dot{x}_i(t_j) = k_i[\sum_{k=1}^{N-1} \eta_k x_{((i+k))_N}(t_j) - x_i(t_j)]$$

with $\eta_k > 0$

$\sum_{k=1}^{N} \eta_k = 1$

where

$k_i$ - Gain for $i^{th}$ agent

$\eta_k$ - Weight assigned to $k^{th}$ agent

$((a))_N$ - $a$ mod $N$

Above system of first order differential equations can be reduced to set of linear equations and motion parameters (Gain, weight) can be evaluated using least-squares on incoming frames. These parameters can then be used for trajectory predictions.

## Results

Videos from surveillance cameras of VMCC, IIT Bombay were used to test the framework. Following shows the results from [2] and studied pipeline using automatic human detection.

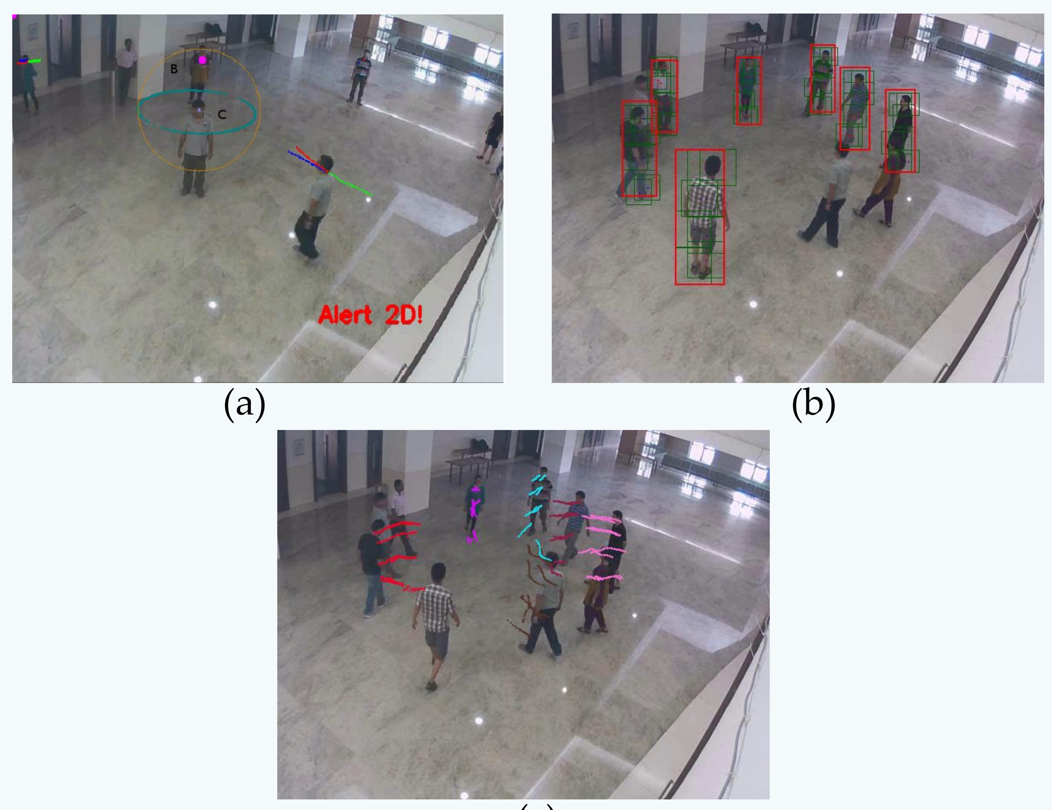



Fig: Results from VMCC dataset
(a) Prediction using LCP (b) Human Detection (c) Tracking using Lucas-Kanade feature points *Fig(a) Courtesy: [2]*

## Conclusion

The pipeline of the algorithm used in the study is similar to [2]. The initial manual labeling has been, however, replaced by automatic detection using deformable part models for intermediate frames.

Further study aims at modeling the behavioral aspects of humans by observing and analyzing various interactions.

**References:**
[1] Discriminatively trained Deformable Part Models, Felzenszwalb et al., 2010
[2] Linear cyclic model based prediction of violation of personal space, Neha et al, 2013
[3] Multi-agent consensus using generalized cyclic pursuit strategies, Arpita Sinha, 2009.