From Crowd Dynamics to Crowd Safety

D. Karthik
13104036

guided by:
Prof. K.S.Venkatesh & Prof. Laxmidhar Behera
Department of Electrical Engineering
IIT Kanpur

June 2016

Outline

Outline

Background

The study of Pedestrian Crowds is an interdisciplinary field, and is carried out within different communities, with different focuses such as:

- **Physicists:** Explaining self-organization phenomena.
- **Fire engineers:** Simulating Evacuation Time of building geometries.
- **Biologists & Sociology:** Investigating how crowd interact with each other.
- **Computer Science:** Large scale multi agent simulations, as well as making realistic crowd animations
- **Computer Vision:** Identification,Tracking,Action Recognition
- **Psychologists:** Interested in the decision-making mechanisms pedestrians are facing.
- **Transportation:** Mobility & flow of crowds and the crowd density.

Project Motivation

- Presently there are large no.of Crowd gatherings all over the world but also there are numerous crowd disasters occurring every year.
- Unfortunately, the information about the (spatio-temporal) development of these events tend to be qualitative rather than a quantitative analysis
- there are not many preventive measures to manage the crowd and control the huge crowd during the critical conditions in order to avoid Crowd Disaster.

Objectives of the Thesis

Purpose: To make a Dynamic Crowd Model and implement the model in real time. So that we can manage the crowd and preventing from **Crowd Disaster**.

.

Problems addressed:

- Head Detection
- Crowd Density
- Crowd Pressure

Outline

Introduction

Aim: To identify the **Location of Head** in the image or frame, which is later used in **crowd modeling**.

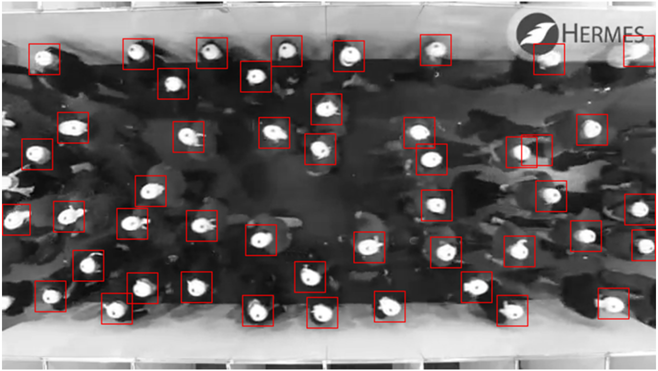


Figure: Location of Head

State-of-the-art

- Crowd Dynamics and Analysis spans wide range of applications such as pedestrians detection, estimating Count of pedestrians, tracking, crowd modeling ,abnormal behavior Detection, Crowd Density Estimation etc.

  .
- Pedestrian detectors focus on **sliding window approaches**. These appear most promising for low to medium resolution settings, under which **segmentation** [1]or **keypoint** [2] [3] based methods often fail.

  .
- **Foreground based methods**:[4] [5] Foreground is extracted firstly by background removal using a reference image, then crowd density is computed as a function of the number of foreground pixels. However, these methods may fail when the background changes gradually over time.

[1]Gu et al. CVPR 2009
[2]leibe et al. CVPR 2005
[3]Seemann et al. BMVC 2005
[4]Xu et al. IEEE ICIP 2005

- Papageorgiou et al. proposed one of the first sliding window detectors, applying support vector machines (SVM) to an over-complete dictionary of **multiscale Haar wavelets**.
- **Viola and Jones** [VJ] [1] built upon these ideas introducing integral images for fast feature computation and a cascade structure for efficient detection and utilizing AdaBoost for automatic feature selection.
- Large gains came with the Dalal and Triggs [HOG] [2] popularized histogram of oriented gradient (HOG) features having gradient-based features. It is inspired by SIFT features [3]for detection which have shown substantial gains.

[1]Viola et al. IJCV 2004
[2]Viola et al. IJCV 2004
[3]Lowe et al. IJCV 2004

HOG Features

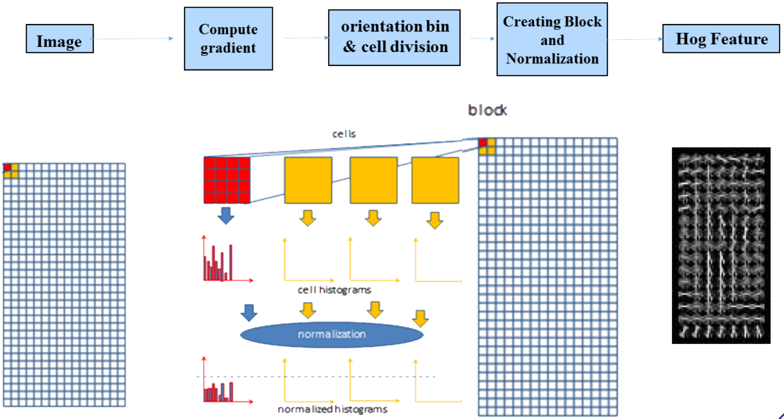


Figure: Histogram of Gradient Feature

Orientation Bins and cells dividing

Gradient image is divided into cells. The size of a cell can be varied. Each pixel within the cell casts a weighted vote for an orientation-based histogram channel based on the values found in the gradient. Histogram channels are evenly spread over 0 to 180 degrees or 0 to 360 degrees, depending on whether the gradient is unsigned or signed. Dalal and Triggs[1]found that unsigned gradients used in conjunction with 9 histogram channels performed best in their experiments.

[1]Dal and Triggs IEEE CVPR 2005

Descriptor Blocks

- The gradient strengths are locally normalized by grouping the cells together into larger, spatially-connected blocks. The HOG descriptor is the vector of normalized cell histograms from all of the block regions.
- These blocks typically overlap. Two main block geometries exist: rectangular R-HOG blocks and circular C-HOG blocks. R-HOG blocks are generally square grids, represented by three parameters: the number of cells per block, the number of pixels per cell, and the number of channels per cell histogram.

Block Normalization

There are different methods for block normalization. Let v be the non-normalized vector containing all histograms in a given block, $\|v_k\|$ be its k-norm for k = 1, 2 and e be some small constant (whose value will not influence the results). Then the normalization factor can be one of the following:

$$L2-norm: \quad f = \frac{v}{\sqrt{\|v\|_2^2 + e^2}}$$

$$L1-norm: \quad f = \frac{v}{\|v\|_1 + e}$$

$$L1-sqrt: \quad f = \sqrt{\frac{v}{\|v\|_1 + e}}$$

Training Hog Features

Training Binary classifier by HOG features We Trained Hog feature by SVM Classifier containing 3023 training and 1228 test heads of 25×25 pixels. The training results show 99.4% accuracy over 1228 test images

Figure: Training Set

Support Vector Machine

Support vector machines (SVMs) is a supervised learning method used for classification, regression and outliers detection.
Consider training set consisting of two classes $\{x^+ \; x^-\}$, then SVM constructs a hyperplane $w^T x + b = 0$. The hyperplane is constructed in such a way that it has maximum margin between two classes.

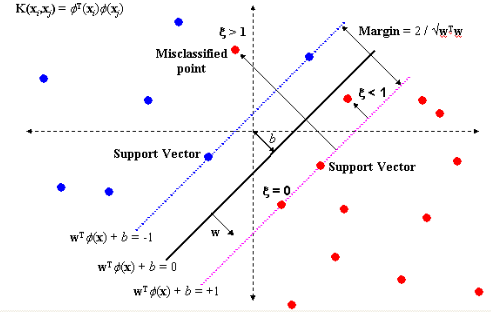
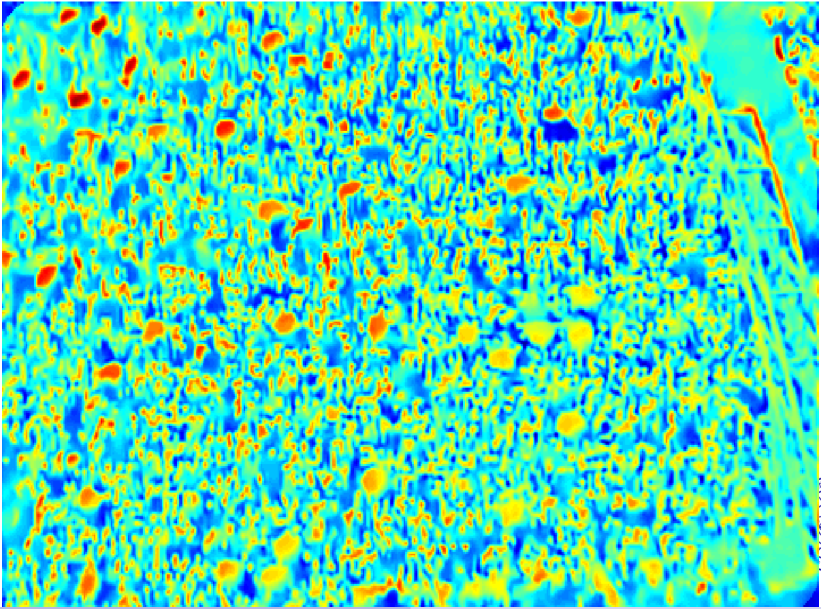


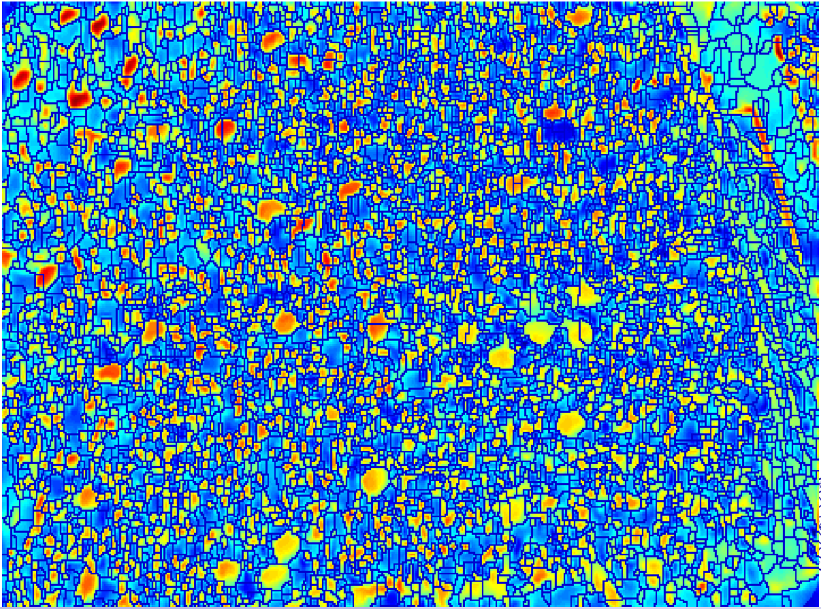
Figure: Linear SVM

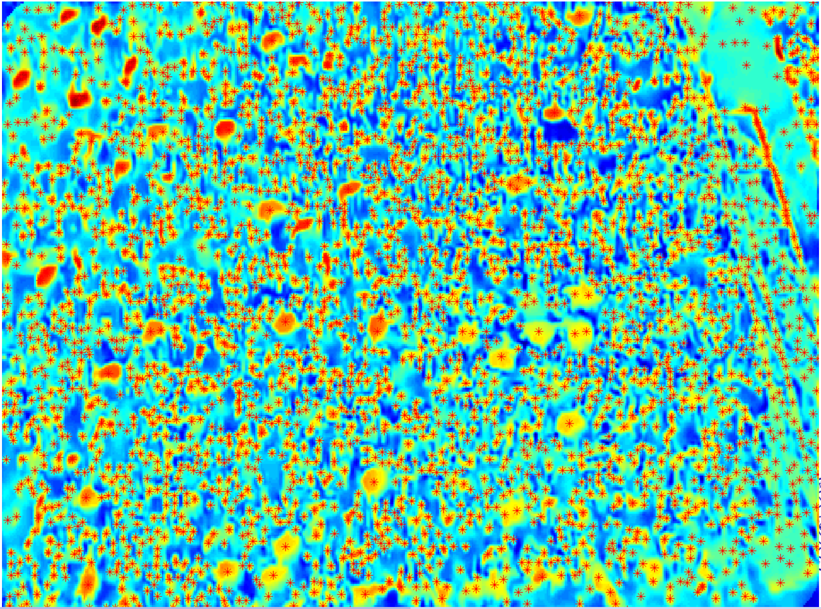
Estimating head by Local Maxima/Local Minima in a image

We find the Local Maxima and Local Minima by finding difference between the median filters where, the size of filter is arbitrary.

Figure: Loc







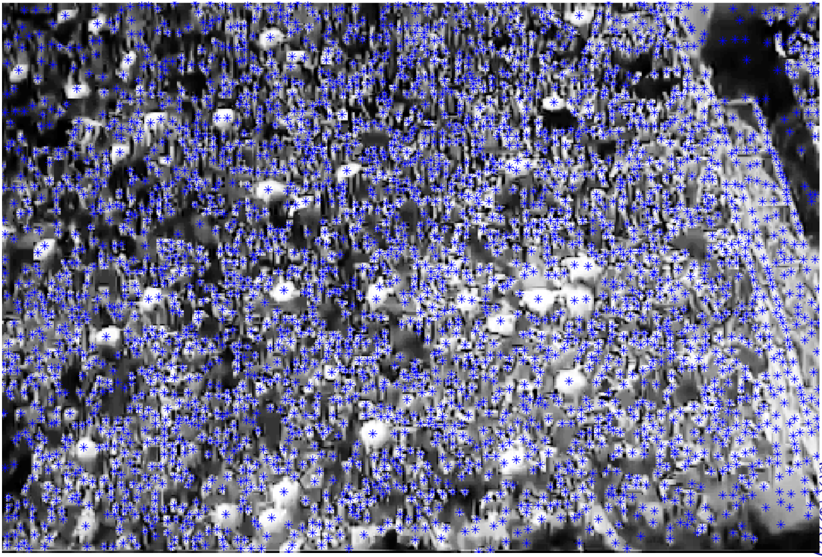


Figure: Frame of Head Detection

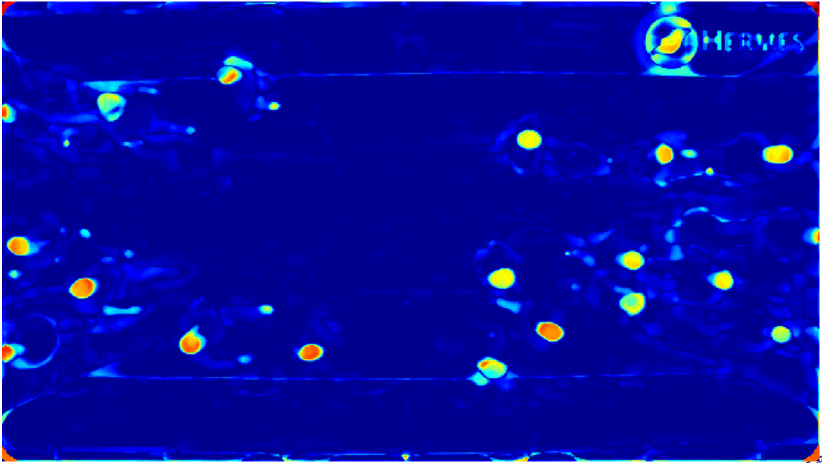
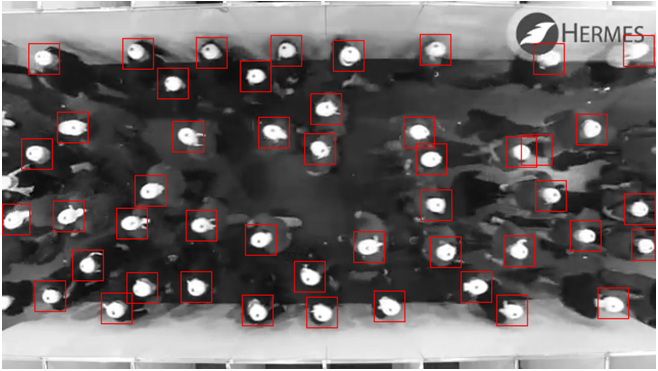


Figure: Frame of Head Detection

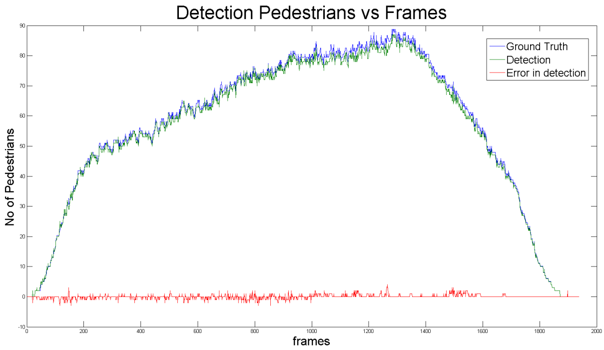
Figure: Local Maxima and Local Minima of Head size

HOG Feature with local maxima/minima

Here we cascasde the input of Local maxima/minima with hog feature and results is as show in figure



Results of Head Detection

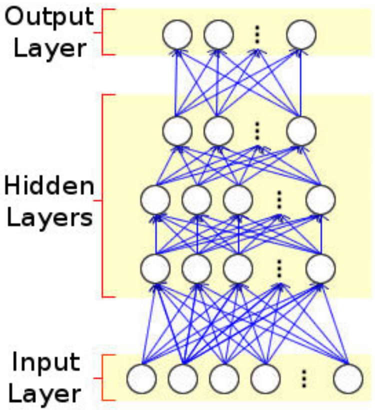


Outline

1. Introduction
2. Head Detection
   - State-of-the-art
   - Extracting HOG Features
   - Estimating head by Local Maxima/Local Minima in a image
   - HOG Feature with local maxima/minima
3. Unsupervised deep learning models for Head detection
   - Deep learning
4. Crowd Modeling

What is Deep Neural Network (DNN) ?

A network with many hidden layers



DNN training using Back-Propagation

- Training of DNN's with conventional Back Propagation algorithm is difficult because of vanishing gradient problem.
- Empirical results show poor performance with more layers.

Solution: Layer-wise Pre-training [Hinton, 2006]

DNN training using Back-Propagation

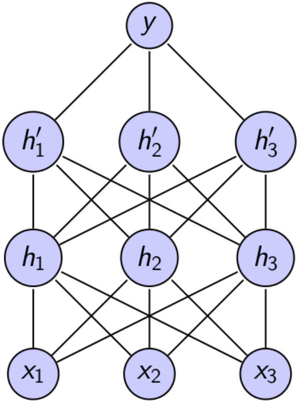
- Training of DNN's with conventional Back Propagation algorithm is difficult because of vanishing gradient problem.
- Empirical results show poor performance with more layers.

Solution: Layer-wise Pre-training [Hinton, 2006]

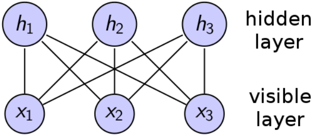
Layer-wise Pre-training and Fine tuning

Unsupervised Learning

- Step 1-i: Train layer 1
- Step 1-ii: Keep layer 1 fixed, train layer 2
- Step 1-iii: Keep layers 1 and 2 fixed, train layer 3
- ... : ... train all layers
- Step 2: Fine tune weights using BP (Supervised learning)



Restricted Boltzmann Machine (RBM)



- x_i, h_j are binary variables $\in \{0, 1\}$
- $x = [x_1\, x_2 \ldots]^\intercal, h = [h_1\, h_2 \ldots]^\intercal$
  - Energy of the model,

$$E(x, h) = -\sum_{ij} w_{ij} x_i h_j - \sum_i d_i x_i - \sum_j b_j h_j$$

  - Probability of model,

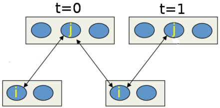
$$p(x, h) = e^{-E(x,h)} / Z$$

  - Normalizer $Z = \sum_{x,h} e^{-E(x,h)}$ is called the partition function
  - Low energy states (x, h) have high probability

Training RBM's: Contrastive Divergence-1

for all m

a. Sample h_j for all j, with fixed $x = x^{(m)}$
b. Sample x_i for all i, with fixed h
c. Sample h_j for all j, with fixed x



$$\langle x_i h_j \rangle_{\text{data}} = \sum_m x_i^{(m)} h_j \quad \text{at } t = 0$$

$$\langle x_i h_j \rangle_{\text{model}} = \sum_m x_i h_j \quad \text{at } t = 1$$

This method is called CD-1

$$\Delta w_{ij} = \eta(\langle x_i h_j \rangle_{\text{data}} - \langle x_i h_j \rangle_{\text{model}}) \tag{1}$$

Experiments

Training is done for 4251 number of 25 $\times$ 25 images broken into 3023 images of class 1(heads) and rest 1228 of class 2 (no head). Then the over-sampled data has been used for further processing by the Deep Belief Network

- Single layer DBN of 100 neurons was constructed initializing all weights and biases to zero.
- It was trained on the full training set, using mini-batches of size 10, with a fixed learning rate of 0.01 for 100 epochs.
- Having pre-trained the weights and biases were used to initialize a neural net with 2 layers of sizes 100 $\times$ 2, the last 2 neurons being the output label units.
- Neural Network was trained using mini-batches of size 2 sample images for 50 epochs using a fixed learning rate of 0.1.
- To evaluate the performance the test set was taken and the maximum output unit was chosen as the label for each sample resulting in an error rate of 0 %. The code ran for 2 minutes.

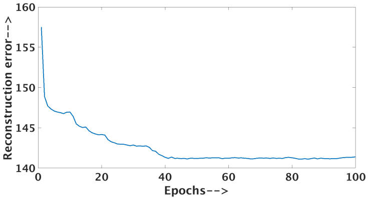


Figure: Reconstruction error

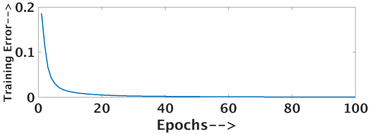


Figure: Training error

Outline

Introduction

To describe the individual and group behaviors in crowded scenes. Many pedestrian models have been developed over the years and have been published. These are basically grouped into two main approaches:[1]

- **Microscopic**
- **Macroscopic**

[1] Johansson et al. Journal of Advances in Complex Systems 2008

Different Approach

Microscopic Approach :The space-time **behaviour of individual pedestrians** rather than describing the behaviour of region or group.
Macroscopic Approach : Macroscopic models focus is to describe the behaviour of group of pedestrians rather than individual pedestrian.

Self-organizing phenomenon :Lanes

Lanes When two streams of oppositely moving pedestrians, intersecting at a 180-degrees angle then pedestrians self organize into **lanes**of uniform direction of motion. This organization of lanes is beneficial for the pedestrians, since it reduces the frequency and strength of avoidance manoeuvers.

Self-organizing phenomenon :Strips

The pedestrians do not only have to avoid collisions with oppositely moving pedestrians, but they eventually also have to traverse the **stream** of oppositely moving pedestrians. Pedestrians with the same desired walking direction turn out to form groups, and the walkway is dynamically subdivided into **stripes** made up by groups with the two different directions of motion.

Self-organizing phenomenon :Bottleneck

During evacuations most narrow places. The flow of pedestrians is proportional to the width and that the flow characteristics are not further changed.

Stop-and-Go Waves

When the density is so high that coordination is difficult a question arises as to whether self-organization survives or will it break down as well ?. We have discovered that, when the density is so high that the level of coordination breaks down, the smooth flow turns into stop-and-go flow [1] For stop-and-go flow, rather than everybody moving at the same speed, the motion is characterized by an alternation of moving and stopping. Basically, the pedestrians are stopped until free space appears in front of them.

[1] Helbing et al. 2007 Physical review E

Crowd Turbulence

The density would reach even higher values and sudden transition from stop-and-go waves to irregular flows starts to take place. These irregular flows were characterized by random, unintended displacements into all possible directions which pushes people around and this is referred to as "crowd turbulence".

This caused some individuals to stumble. As the people behind were moved by the crowd as well and could not stop, the fallen individuals were trampled, if they did not get back on their feet quickly enough. Tragically, the area of trampled people grew more and more in the course of time, as the pedestrian became obstacles for others which lead to biggest crowd disasters.

Crowd Model

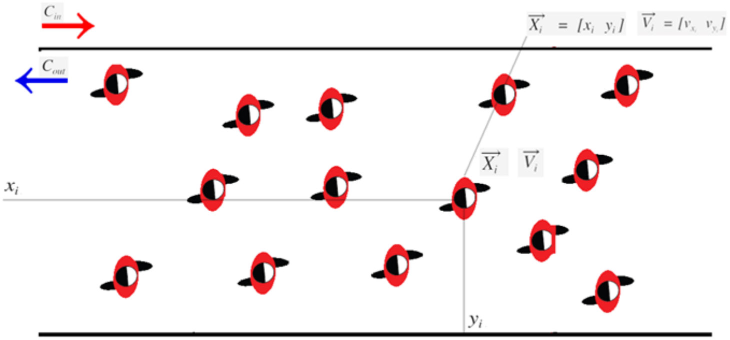


Figure: Crowd Model

View Point Variation of Video Dataset

Different angles of camera

- **Perpendicular View**
- **Perspective View**
- **High Altitude View**

Perpendicular View

The camera is placed perpendicular to the ground where crowd gathering is going to take place.
The size of head is same in entire Video dataset
No occluded of pedestrians.

Projective View

The camera is placed faraway from crowd and angle is not perpendicular to the ground.
The size of head is not same in the entire video dataset, the person who are near camera has large size. As the angle of camera increases, the chances of occlusion also increases and
at a certain angle the pedestrian is occluded for such type of video dataset. We can$'t$ implement our model on the view as

High Altitude View

The camera is placed at a very large height (an aerial view of the crowd).
If an individual person in the video is very small, say almost 4-5 pixels.

Figure: High Altitude View of Camera

Local Density

Introduction Traditional definition of crowd density is given as no. of person$'s$ in 1 meter square normally. But this is only valid for a uniform pedestrian distribution in the area. The **Average Density** or **Global Density** is given as

$$\varrho = \frac{N_r}{A_r} \tag{2}$$

Note: Global Density Does not give information where the density is high or low in given location.

Vornoic Distribution based Local Density :(ρ_V)

The location of pedestrian X_i is considered as a nucleus and the entire region(A_r) is divided into individual regions of vornoic cells V_i of area a_i.

$$\rho(V_i) = \frac{1}{a_i} \tag{3}$$

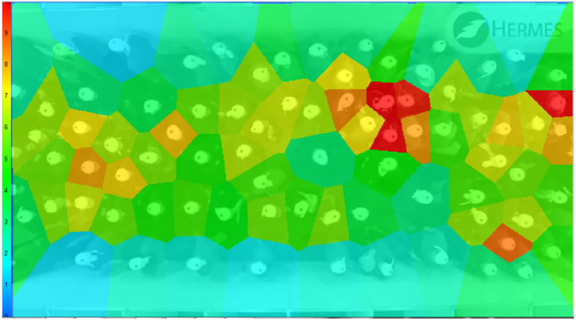


Figure: Vornoic Distribution Based Local Density

Vornoic Distribution based Local Density :(ρ_V)

We can calculate the average of the local density

$$
\begin{aligned}
avg(\rho_V) &= \frac{\iint_{A_r} \rho_V \, dx \, dy}{\iint_{A_r} \, dx \, dy} \\
&= \frac{\sum_j \iint_{a_j} \rho_V(V_j) \, dx \, dy}{A_r} \\
&= \frac{\sum_j \iint_{a_j} \frac{1}{a_i} \, dx \, dy}{A_r} \\
&= \frac{N_r}{A_r} \quad = \varrho
\end{aligned}
\tag{4}
$$

From above derivation, we can find the global density ϱ.

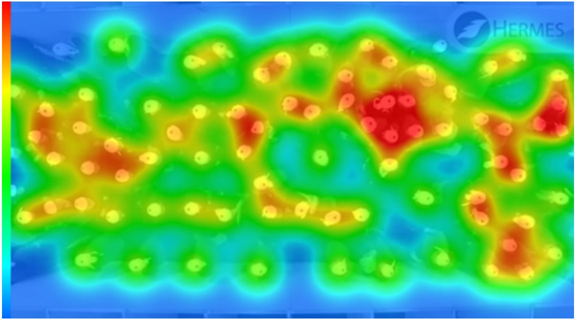
$$
\varrho = avg(\rho_V)) \tag{5}
$$

Gaussian Distribution based Local Density (ρ_G)

The location of i^{th} pedestrian$\overrightarrow{X_i}$ we assume density is gaussian distributed of having a peak of 1 person per $meter^2$ and having standard deviation σ.

$$\rho_G(\overrightarrow{X}) = \quad \frac{1}{\sqrt{2\pi}\sigma} \sum_j e^{\frac{-||\overrightarrow{X} - \overrightarrow{X_i}||^2}{2\sigma^2}} \tag{6}$$

where, standard deviation σ is an arbitrary value



The global density $\varrho(t)$ for different values of σ evaluated over average of local densityρ can be given as:

$$\begin{aligned} avg(\rho_G) &= \frac{1}{\sqrt{2\pi}\sigma} \frac{\sum_j \iint_{A_r} e^{\frac{-||\overrightarrow{X}-\overrightarrow{X_j}||^2}{2\sigma^2}} dx\, dy}{A_r} \\ &= \frac{1}{\sqrt{2\pi}\sigma} \frac{\sum_j \sqrt{2\pi}\sigma}{A_r} = \frac{N_r}{A_r} = \varrho \end{aligned} \tag{7}$$

From above derivation, one can find the global density by:

$$\varrho = avg(\rho_G) \tag{8}$$

From above, ϱ is independent of σ but ρ_G in dependent on σ

Estimating σ

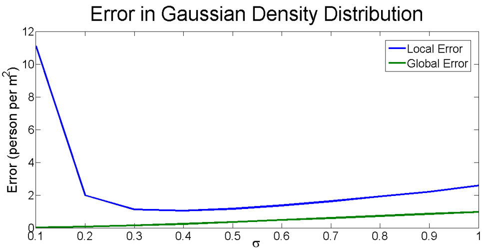
Since ρ_G in dependent on σ. In order to choose desired value of σ we plot the error in ϱ & ρ_G for σ value 0 to 1.

$$\textbf{Global} \quad \textbf{Error}(\sigma) = \sqrt{\frac{1}{F} \sum_{f=1}^{F} [\ \varrho(f) - avg(\ \rho_G(\sigma, f)\ )]^2} \qquad (9)$$

where f is current frame and F is total no of frames.

$$\textbf{Local} \quad \textbf{Error}(\sigma) \ = \sqrt{\frac{1}{F} \sum_{f=1}^{F} [avg\,(\rho_V(f)) - \ \rho_G(\sigma, f))]^2} \qquad (10)$$

f


Error in Gaussian Density Distribution
Local Error
Global Error
Error (person per m2)
σ
0
2
4
6
8
10
12
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Results of Local Density

In order to find the local density ρ of the given image we need to find the head location $X_i(t)$ which is calculated by using HOG features and results of global density for all frames is given in figure.

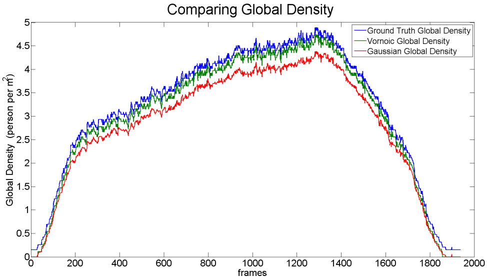
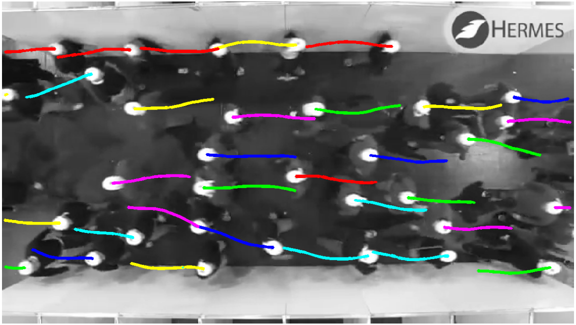


Figure: Global Density Vs frame

Local Velocity

Velocity Distribution of the region(A_r) can be considered to be either a gaussian or a vornoic distribution of the individual pedestrian X_i.
In the case of Microscopic Approach where number of pedestrians is not more than few hundred, we use kalman filter tracking to results of kalman tracking is given in figure **??** to find the velocity of individuals V_i.



Optical flow based Local Velocity estimation

In Macroscopic Approach in such a case, the number of pedestrians are very large (in thousands) and the area occupied by a pedestrian is very small(few pixels), hence detection of individual person is very difficult and applying tracking algorithm will be much more complex and non-approachable but we can approximate velocity of individual person as the optical flow. However this is only valid if the pedestrian occupies very small area(few pixels) but if the pedestrian occupied area is large, this approach gives unwanted velocity and further we can't find the change in velocity by optical flow for large pedestrian size.

Variation in Velocity

The individual pedestrian behaviour and velocity most importantly depends upon the surrounding behaviour and velocity of pedestrians rather than the previous velocity of the individual pedestrians

Then variation in velocity $\overrightarrow{v}$ is given as

$$\nabla v = v - avg(v) \tag{11}$$

For Macroscopic Approach we can consider the above parameters (pressure, velocity, density) to be continuous in space $\overrightarrow{X}$ which means at each pixel considered, an individual person has his/her own velocity υ. Then variation in velocity $\overrightarrow{\upsilon}$ is given as

$$\nabla \upsilon = \upsilon - \upsilon * N\left(0, \sigma^2\right) \tag{12}$$

Note: Only valid only when we assume the adjacent pixel is also a pedestrian having velocity value and υ is calculated either by individual velocity or Optical flow based Local Velocity.

For microscopic approach : the distribution is mostly discrete, so for the variation in velocity of the i^{th} pedestrian we use a similar approach as above and avg(υ) is given as gaussian average of neighboring pedestrians with $\sigma = 0.4$ and given as

$$avg\left(\upsilon_i\right) = \frac{\sum_j \left(\upsilon_j\right).e^{\frac{-\left\|x_j - x_i\right\|^2}{2\sigma^2}}}{\sum_j e^{\frac{-\left\|x_j - x_i\right\|^2}{2\sigma^2}}} \tag{13}$$

and the Variation in the velocity of i^{th} is given as

$$\nabla\upsilon_i = \upsilon_i - avg(\upsilon_i) \tag{14}$$

Pressure

In order to measure the irregularities in crowd which is also refereed as crowd turbulence, we use information about density and change in velocity.
The Crowd Pressure is given as:

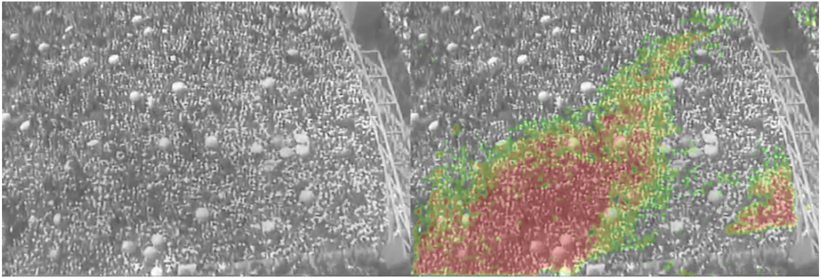
$$P(\overrightarrow{X}_i) = \rho_i \nabla \overrightarrow{v_i} \tag{15}$$

Similar to the Newtonian Pressure where we consider area of the individual as crowd density and mass of the individual is considered as a unit value.

$$\begin{aligned} P(\overrightarrow{X}_i) &= \frac{1}{A_i} F_i \\ P(\overrightarrow{X}_i) &= \rho_i * 1 * \nabla \overrightarrow{v_i} \\ P(\overrightarrow{X}_i) &= \rho_i \nabla \overrightarrow{v_i} \end{aligned} \tag{16}$$

where ∇v_i is referred as variation in velocity but not change of velocity with respect to time but with respect to $\overrightarrow{X}$.

We have evaluated the pressure parameter values and the visual results of pressure changing in a given image as shown in figure and the pressure variation for overall frame. The results are similar to the results produced by [2]



[2] Helbing 2007 Physical Review E

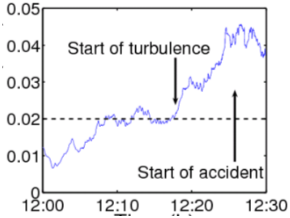


Figure: reference Pressure Vs frame

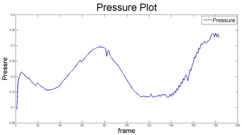


Figure: Pressure vs frame

Outline

Conclusion

The main contributions of the present work are as follows:
In the 2^{nd} chapter,

- We use novel Local Maxima/Local Minima method in order to estimate head in an image, where we use simple Difference in median filter in order to estimate the Head.
- The major contribution to the work is by combining hog with local maxima and minima algorithm where it reduces the computation time, hence paving a way for implementation of the crowd model in real time.

In 3^{rd} chapter

- Unsupervised deep learning models have been used for detecting heads and achieving close to 100% accuracy.

Conclusion

In 4^{th} chapter

- We have tried to implement the crowd model for different camera angles such as perpendicular view, perspective view, high altitude view even though there are limitations in the implementation.
- We also theoretically redefine the Gaussian Local Density based on Helbing et al. 2007 , which is independent of camera angle and height of camera unlike traditional approaches in Johansson et al. 2007and it is more realistic than traditional way of definition.
- We normalized the Gaussian Local Density which is independent of the camera angle and height. We evaluated best sigma ($\sigma = 0.4$) value suitable for the Gaussian Local Density.
- We have redefined the variation in velocity which can be used for both microscopic and macroscopic approaches.
- We have implemented a novel, simple way of calculating the variation in velocity by using optical flow as velocity and the results are similar to the traditional methods of calculating variation in velocity.

Future Scope

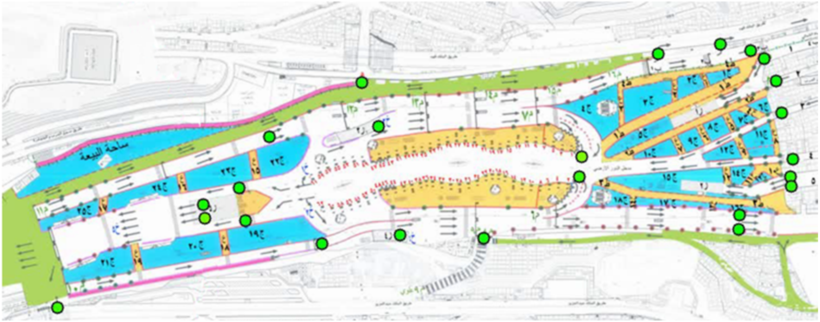


Figure: reference Pressure Vs frame

Future Scope

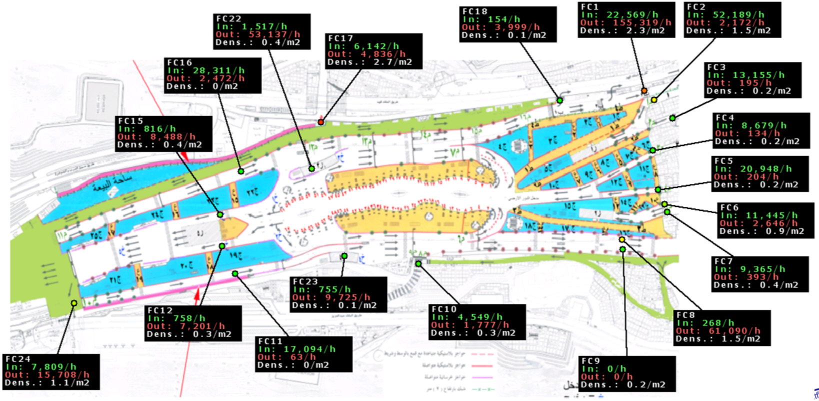


Figure: reference Pressure Vs frame