ECE/CSE 478 Digital Image Processing

Multi-Source Approach for Crowd Density Estimation in Still Image

Introduction

Estimation of people density in intensely dense crowded scenes is very crucial due to perspective difference, few pixels per target, clutter and complex backgrounds etc. Mostof the existing work is unable to handle the crowds of hundreds or thousands. At this level of density, one feature is not enough to estimate the total density of an image. We propose a hybrid model which relies on multiple source of information as Fourier analysis, Local binary pattern, Gray level dependence matrix (GLDM) features and Histogram of oriented gradient (HOG) for head detection to estimate the total count. Each of these features separately contribute in final total count estimation along with other statistical measures. Experiential results validate the performance of our proposed approach by computing the total count with respect to ground truths.

Crowd analysis and monitoring is an essential task in public places to provide safe and secure environments. In past decades, many crowd disasters had taken place due to lack of crowd control management. Examples of crowd tragedies at mass gathering events include Mina-Mecca stampede in 2015, Boston marathon bombing in 2013 and Love Parade disaster at Germany in 2010.


Fig1 &2: Depicting Boston Rally and Kumbhmela Stampede

Motivation

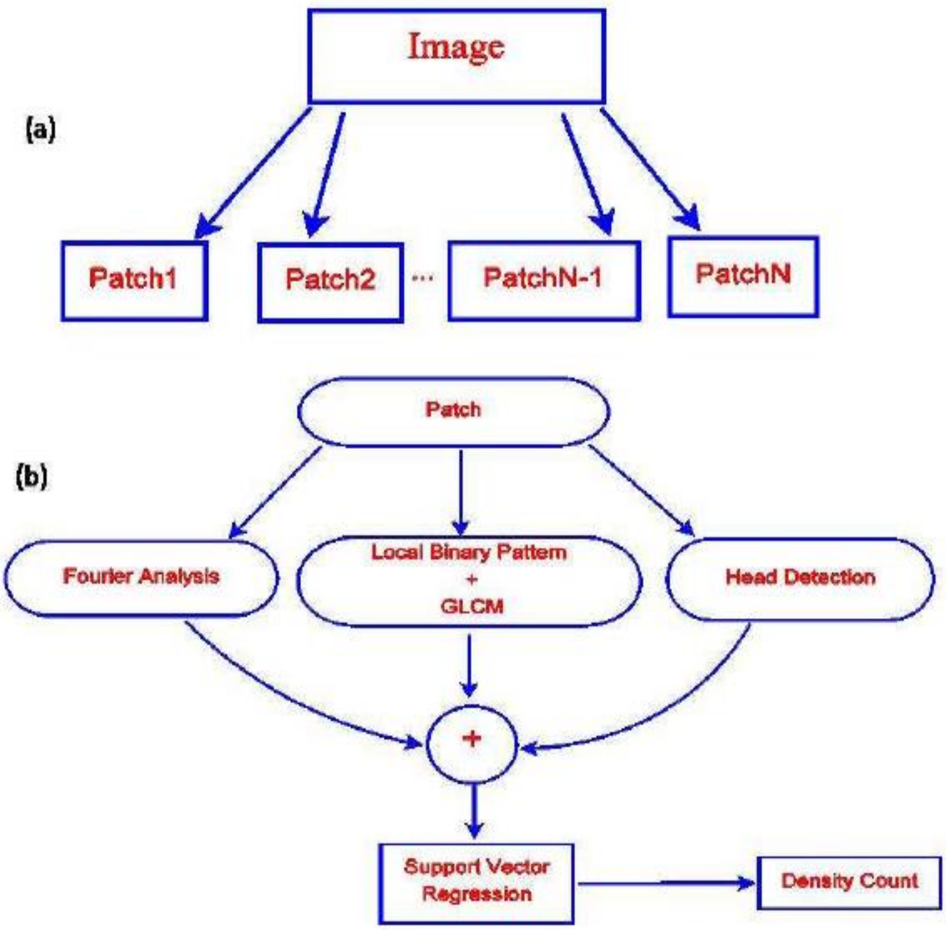
Appropriate crowd control and its management is required to avoid crowd related disasters caused by large gathering of individuals and generate a quicker response to the emergent situations. Crowd density is one of the primitive explanation of crowd status. Counting the number of individuals include several real-world applications as crowd management, design of public spaces and safety [1] and security services. People counting or density estimation is a vital task in many crowded scenarios such as public rallies, religious gatherings, sports stadiums, and transportation stations.

Problem Statement / Real Issue

Not one method is perfect for solving the problem ,hence a combination of all Fourier Analysis ,Hough Transform &LBP is used . The significant addition of the proposed approach is the utilization of frequency-domain, feature extraction ,and texture analysis method in crowd density estimation. Fourier transform has been extensively used in texture analysis .

In a reliable estimation of the texton counts is provided by Fourier transformation. Application of frequency domain analysis is limited because of irregularity in spatial arrangement of texture elements and failure in localization of repeated elements. To overcome these drawbacks, we present an innovative idea. We employ Fourier analysis along with Histogram of Oriented Gradient part based head detection and local binary pattern based gray level co-occurrence matrix to ignore the problem of irregularity in the perceived textures coming from images of dense crowded scenes. The individual count is estimated by individual approach and then aggregated subject to global consistency constraints.

Method of Approach

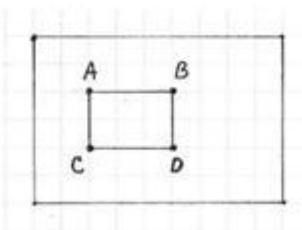


In the following session , each method of approach can be seen .

Additional Method Approaches

We can use other effective Voila-Jones Algorithm for head detection , in this algorithm the head detection is much faster and is more efficient .The main property of this algorithm is that training is slow, but detection is fast. This algorithm uses Haar basis feature filters, so it does not use multiplications. The efficiency of the Viola-Jones algorithm can be significantly increased by first generating the integral image.

$$II(y,x) = \sum_{p=0}^{y} \sum_{q=0}^{x} Y(p,q)$$



The integral image allows integrals for the Haar extractors to be calculated by adding only four numbers. For example, the image integral of area ABCD (Fig.1) is calculated as

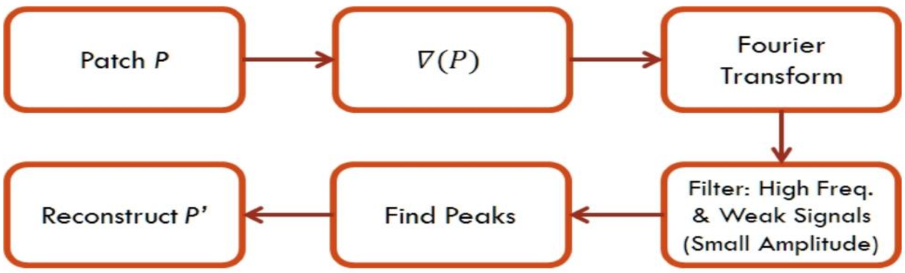
Area= $II(y_A,x_A) - II(y_B,x_B) - II(y_C,x_C) + II(yd,xd,)$.

Detection happens inside a detection window. A minimum and maximum window size is chosen, and for each size a sliding step size is chosen. This decreases the time complexity and computing complexity too.

Fourier Analysis

In crowd images, when thousands of people are present, each of people occupy only few pixels per target and some people are very far away from the camera with perspective distortion. They have monotonous texture, density of crowd in patch can be uniform and can be captured by Fourier Transform (f), where the sporadic occurrence of human heads shows as peaks in the frequency domain.

Algorithm for Fourier Transform :



Now , for inputting the data into the SVR –Support Vector Regressor , we need to find various features like : Kurtosis, Variance , Entropy & Skewness. These parameters characterise the image uniquely and hence can be used as a part of training data . These parameters are calculated both for Patch as well as |P-grad(P)| for more data .

Data set : UCF data

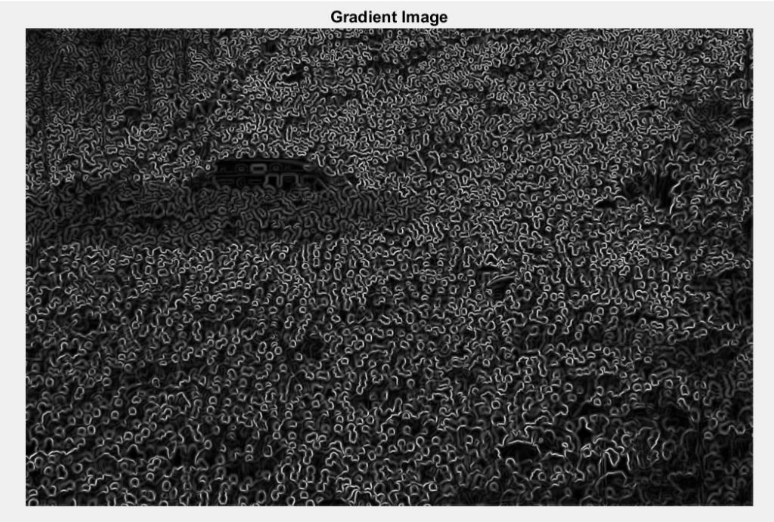
Disadvantage of Fourier Analysis :

1. At places ,where there is discontinuity it can't be efficiently implemented .

Stage-wise Output Images

Input Image :

Gradient of the Image : Sobel Filtering



Fourier Transform of the Gradient image :

Gaussian Low Pass Filtering

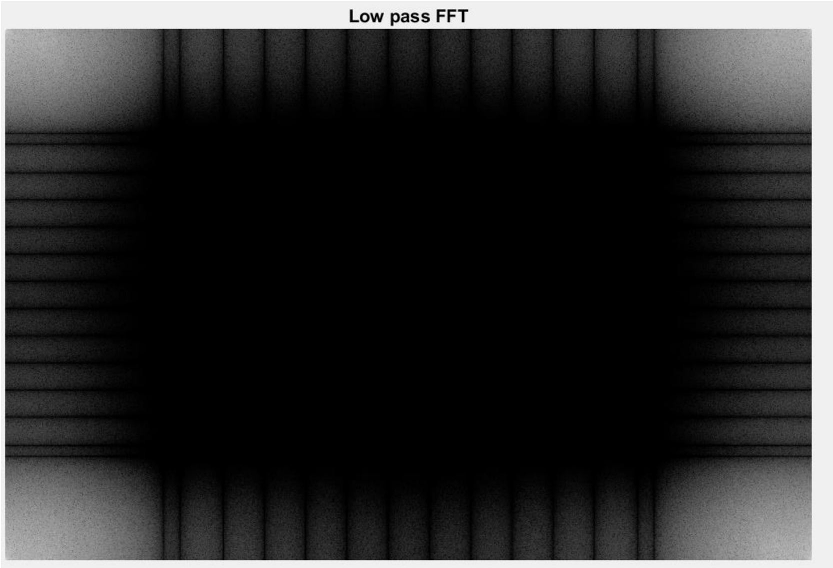
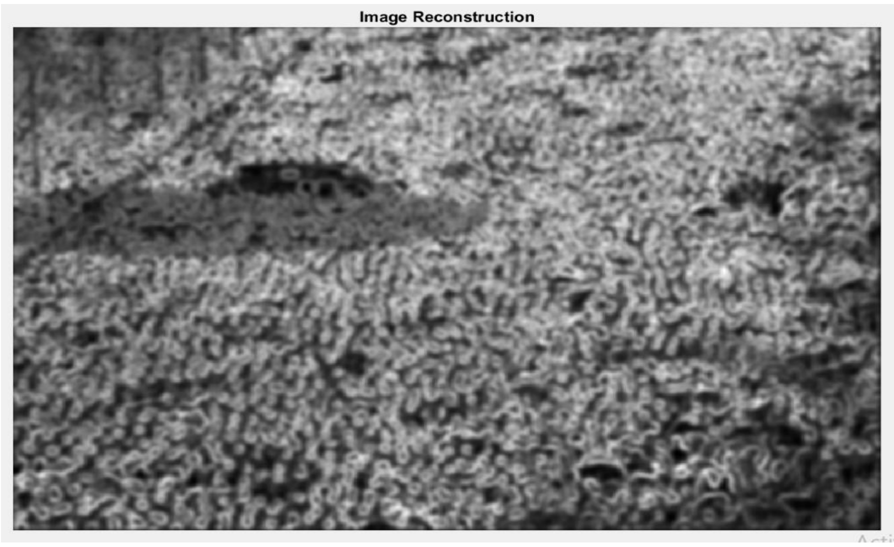
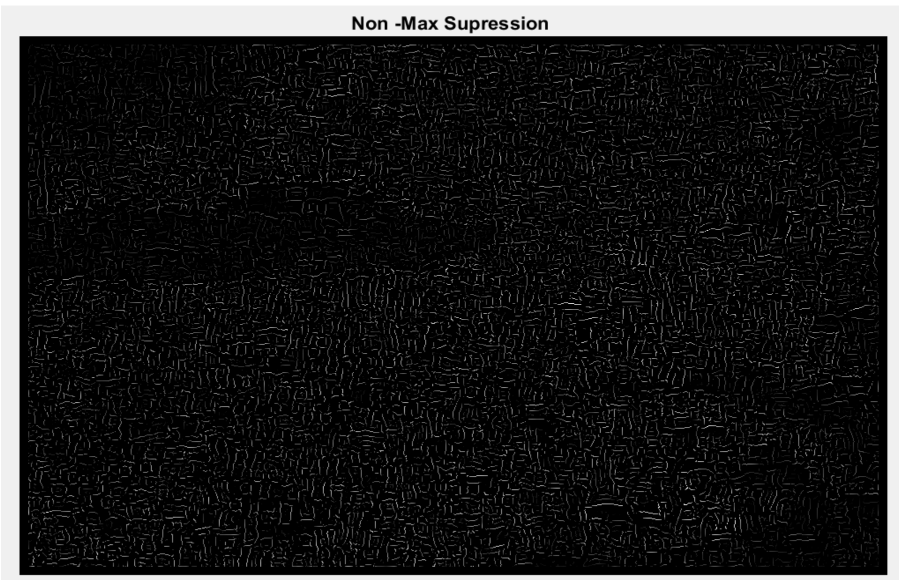


Image Patch Reconstruction (From Gradient of the image)



Non –Maximal Suppression :



Output Feature Vector :

Calculations :

$$\text{Energy}(d,\theta) = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} f(i,j|d,\theta)^2$$

$$\text{Contrast}(d,\theta) = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} (i-j)^2 f(i,j|d,\theta)$$

$$\text{Homogeneity}(d,\theta) = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} \frac{f(i,j|d,\theta)}{1+(i-j)^2}$$

$$\text{Entropy}(d,\theta) = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} f(i,j|d,\theta)$$

[0.500 0.8442 -0.0010 0.0114] = { Energy , Variance, Kurtosis , Skewness } .

Hence , this same procedure has to be done on each patch and supplied to the SVR Machine .

We can also compute the same for the difference image which will be the difference of that patch and the gradient .

Support Vector Regression – SVR

The feature vectors relating to each of analyiss –HoG ,LBP+GLCM &Fourier analysis were implemented by using Image Block Processor , and then the text files are fed as input to train the model . The output is the count (estimate) for the new image that the SVM will take.

The trained model has taken 7 input feature vectures , and over 50 images.

The dataset used here :UCF 50 Data Set

Attached files : CC2.txt, FAD.txt,varhog.txt & meanlbp.txt

Sample Output :

- RESULTS :

SVM Estimator :
1208 counts

Ground Truth :
740

Analysis & Discussion :

1. Problems in picking the right kind of images .
2. Patch division in each code.
3. SVR regression results-Patches issue

Failure Cases :

Images that contain building / statues/etc are also falsely detected . The error rates the code is high. In one of the cases , a shadow of the image had fallen on the crowd , which lead to black patch in the output image analyis , which further decreased the count of the total image .

GitHub Repository Link

Link : https://github.com/NiharikaMessi/Project-DIP-

Task Table:

1.LBP + HOG Coding = Deepa .P

2.Fourier Analysis = Niharika

4.SVR Coding &Implementation = Niharika

3. Report – Niharika (60%) & Deepa- (40%).

4.PPT- Niharika &Deepa

-END-