

Crowd Analysis and Density Estimation Using Surveillance Cameras

Nachiket Kulkarni , Abhijeet Rana, Aboli Patre
P.E.S. Modern College of Engineering, Pune

Abstract-This paper aims at describing an automated process for crowd analysis using surveillance cameras. Taking automation of crowd analysis into consideration, we present a methodology for creation of automated crowd analysis systems and its implementation in organizations like shopping malls and super markets.

The purpose of this paper is to present an automated system that would estimate the number of people in a particular region of interest and also help us understand the crowd distribution by generating graphical reports based on the surveillance data. We also aim to provide further assistance to our customers in the form of suggestions derived from the crowd analysis that would help them to take effective business decisions.

General Terms

Crowd, Surveillance, Analysis

Keywords

Crowd Analysis, Image Processing, Density Estimation, Report generation.

1. INTRODUCTION

CCTV (Closed Circuit Television) are used as a traditional approach to monitor crowd. Although CCTV surveillance has proved to be an effective technique to monitor crowd, it takes human visual ability for the same. The present system is not fully automated. There is a need of automated system to estimate size or density of crowd.

Crowd Analysis is the process of understanding the behavioral patterns and density of a particular cluster of a crowd. Using crowd analysis we can determine the number or type of people at a certain location. The density of the crowd and their behavioral patterns can be analyzed and understood. This is done by continuous or periodic monitoring of the crowd using cameras.

Crowd density estimation methods can be divided into four categories –

(1) Background Removal Technology

It is an image processing method based on pixel statistics, calculating the number of foreground pixels through background subtraction technique, and estimating crowd density by the number of pixels. Although this method is simple and effective, it is ineffective for high density crowd.

(2) Pattern Recognition Technology

It presents the crowd density estimation method of texture analysis. The advantage of using this method is that it solves the problem of overlapping, so that high density crowd can be estimated by this method. However, the

disadvantage is that the accuracy is very low for low density crowd.

(3) Information Fusion

The kalman filtering method can count number of people by the background removal technology and boundary technology.

(4) Combination of Pixel Statistics and Texture Analysis

This method combines the advantages of both pixel statistics and texture analysis.

The proposed method by Ming Jiang and Jingcheng Huang [1] reduces the impact of perspective distortion by dividing the region of interest. This method can improve the accuracy in the whole density range. Also, estimating the crowd size for the high density and extremely high density crowd. Their experimental results show that the proposed method can improve the accuracy in the whole density range.

2. SURVEY

The current system is based on only crowd monitoring but not crowd analysis. Behavioral patterns of crowd and their densities are not regarded as essential information by developers of current systems. The current system does not incorporate a central database storing information about the crowd visiting a store or region. It simply stores sales amounts.

Current systems does not provide precise information on which decisions can be taken to optimize the business strategies. Current systems are not based on intelligent monitoring of crowd and do not determine the relation between crowd distribution and the business efficiently. A survey has been done in paper "Counting people in the crowd using a Generic Head Detector" [2] that most of the systems use Head region to detected human since it is the most visible part of the body in a crowded scene. Only head is visible thus various head detector algorithms are used for the same.

Also the paper - "An accurate algorithm for head detection based on XYZ and HSV hair and skin color models" [3] suggests to use a new model for the head detection based on appearance distributions and shape constraints. The appearance distribution models the colors of hair and skin by sets of Gaussian mixtures in the XYZ and HSV color spaces. This is done as it is not necessary that the human head or face will face the cameras all the time. It may also

happen that humans face the opposite side of the cameras, thus for which the system needs to be trained for detecting human hair as well . It is considered that the human face is in an elliptical shape, which makes it easy for the system to distinguish head from rest of the body. Another paper-”An Approach for Crowd Density and Crowd Size Estimation” [4] proposes a new approach of crowd density estimation. This method combines the advantage of pixel statistical feature and texture analysis, and reduces the impact of perspective distortion by dividing the region of interest. Moreover, it estimates the crowd size for high density and extremely high density crowd.

3. PROPOSED APPROACH

We propose an automated system for a shopping mall or supermarket that monitors the crowd and estimates its density. For which we refer [4]. Combining the advantages of Pixel Statistics feature and Texture Analysis, reduces distortion.

Generally an image’s regions of interest will focus on objects in its fore- ground.

After the stage of image preprocessing object localization is required which may make use of this technique. Background subtraction is a widely used approach for detecting moving objects in videos from static cameras. The rationale in the approach is that of detecting the moving objects from the difference between the current frame and a reference frame, often called background image, or background model.

Background subtraction is mostly done if the image in question is a part of a video stream. Background subtraction provides important cues for numerous applications in computer vision, for example surveillance tracking or human poses estimation.

However, background subtraction is generally based on a static background hypothesis which is often not applicable in real environments. With indoor scenes, reflections or animated images on screens lead to background changes. In a same way, due to wind, rain or illumination changes brought by weather, static back- grounds methods have difficulties with outdoor scenes.

3.2 Gaussian Mixture Model

A Bayesian Gaussian mixture model is commonly extended to fit a vector of unknown parameters (denoted in bold), or multivariate normal distributions. In a multivariate distribution (i.e. one mod- elling a vector x with N random variables) one may model a vector of parameters (such as several observations of a signal or patches within an image) using a Gaussian mixture model prior distribution on the vector of estimates given by

$$p(\theta) = \sum_{i=1}^K \phi_i \mathcal{N}(\mu_i, \Sigma_i)$$

where the i 'th vector component is characterized by normal distributions with weights ϕ_i , means μ_i and covariance matrices Σ_i . To incorporate this prior into a Bayesian

estimation, the prior is multi- plied with the known distribution $p(x|\theta)$ of the data x conditioned on the parameters θ to be estimated. With this formulation, the posterior distribution $p(\theta|x)$ is also a Gaussian mixture model of the form

$$p(\theta|x) = \sum_{i=1}^K \tilde{\phi}_i \mathcal{N}(\tilde{\mu}_i, \tilde{\Sigma}_i)$$



Fig. 1. Sample Area



Fig. 2. Crowd count

3.1 Background Subtraction

Background subtraction, also known as Foreground Detection, is a technique in the fields of image processing and computer vision wherein an image’s foreground is extracted for further processing

3.3 Pixel Statistics feature

The property of the pixel statistic is the earliest feature to be used for crowd density estimation, and it is a very effective feature. The basic idea of this algorithm is denser the crowd, greater the proportion of the foreground image. Researchers considered that there is a linear relationship between the number of foreground pixel and number of people in the scene. Pixel features usually are the foreground image area, perimeter, edge pixels and so on.

Pixel statistical algorithm is relatively easy to understand, has low computational complexity. The relationship between the number of people and pixel feature is relatively simple after preprocessed, easy to train, and the generalization ability of classifier or function relationship is very well after training. However, the pixel statisti- cal algorithm has some problems. Foreground image segmentation algorithm is not ideal for high density

3.4 Texture Analysis

The pixel is very important feature among crowd density estimation, but the accuracy is very low for more serious occlusion area. To solve this problem, Marana proposed texture analysis algorithm. Different density crowd has different texture pattern for texture analysis. Images of low density crowds show coarse texture, while images of high density crowds show fine texture. The calculation of GLCM texture features is a common and effective method.[4] proposes texture method to estimate the crowd density of extremely high density.

Steps to be followed -

- (1) Capture periodic snapshots from CCTV
- (2) Extract foreground image
- (3) Divide image into regions of interest
- (4) Count number of people in each regions
- (5) Aggregate count of all the regions
- (6) Estimate number of total people in the image.
- (7) Data at the database will be aggregated according to the various criteria required for decision making and report generation.
- (8) Reports and graphs along with suggestions are presented to the business owner on the basis of which he or she can make appropriate and best decisions to optimize the business strategy.

The snapshot image is divide into Regions of Interest, which pro- vides the flexibility to estimate the count of the regions whose count of human matters to us. Due to this we can estimate the count that is useful us and eliminate the humans that are irrelevant.

Pixel Statistics feature can estimate count of human in less crowded place very accurately. While Texture Analysis can estimate count of human in heavily crowded place very accurately. If the place is average crowded combining both Pixel Statistics feature and Texture Analysis can provide us with good results as compared to other algorithms.

To address the issue of crowd analysis, a automated system will be designed, implemented and evaluated of the crowd analysis sys- tem. Under this system, an area like a shopping store will be monitored both continuously and periodically using cameras. The data obtained from the image processing of snapshots and footage will be stored, aggregated and reports will be generated.

These reports will not just be presentation of information from the surveillance but also an aid to make vital decisions that will optimise the business strategy. The purpose of this system is to provide an automated way to analyze crowd in an area and present raw data in a meaningful and understandable way.

4. PROPOSED FLOWCHART

The figure below represents the flowchart of the proposed system.

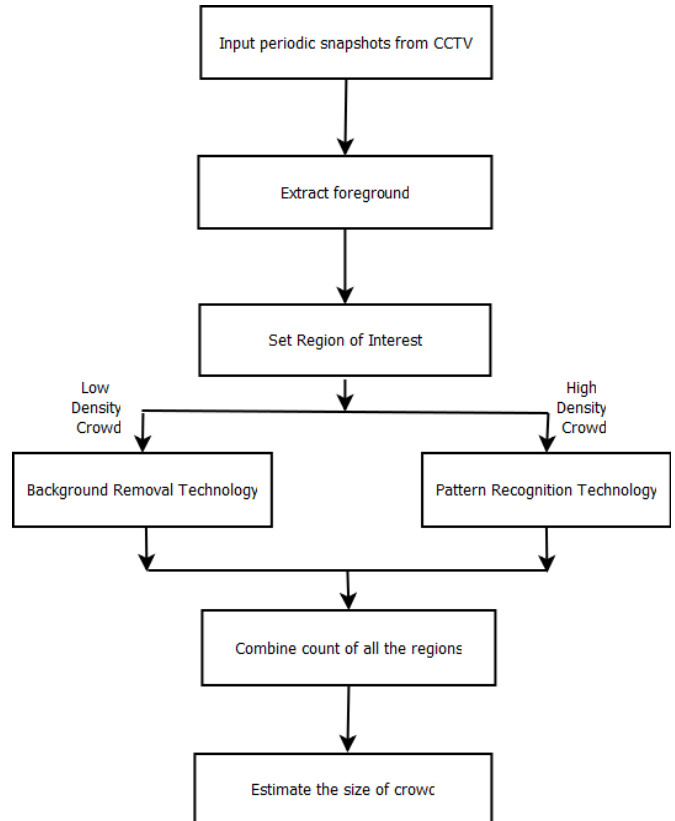


Fig. 3. Proposed Flowchart of the system

5. RELEVANT MATHEMATICAL MODEL

$S = \{ I, P, A, F, R \}$

Input $I =$ Set of Snapshots taken by camera $C C = \{ c1, c2, c3, \dots \}$

Processing $P =$ Set of processed images by client $U U = \{ u1, u2, u3, \dots \}$

Algorithm $A = \{$ Set of algorithms $A1, A2, A3, A4 \}$

—A1 - Algorithm to extract head count from each snapshot using

Gaussian Mixture Model.

—A2 - Algorithm to aggregate data based on frequency and criteria using Apriori.

—A3 - User defined algorithm to create recommendations and sug- gestions using user defined algorithms.

—A4 - Algorithm to generate reports based on aggregated data.

Function $F = \{$ Set of Functions $F1, F2, F3, F4, F5, F6 \}$

—F1- Function to process image and calculate head count using

Gaussian Mixture Model and pixel texture analysis

—F2- Function to send head count from $u1$ where $u1 \in U$ to server using head counting algorithm

—F3- Function to aggregate data using Apriori

—F4- Function to search data based on criteria

—F5- Function to recommend strategies and give suggestions

—F6- Function to generate reports, graphs

Output R = { Set of Output { Density, Reports , Graphs , Suggestions , Recommendations } }

6. ESTIMATION OF CROWD

The image would be divide into several regions, each region having certain number on people into it. Count of each region will be calculated according to its property of less crowded, average crowded and heavily crowded place. Depending on which the algorithms will be applied on the regions.

After each regions density is crowded, aggregation of count of all regions will be done and the total crowd density can then be estimated accurately. It would give of how many total number of humans are there in the image.

Function Mappings :

- F1(h) —> R(Density)
- F2(x), F3(x), F4(x) —> R(Density)
- F5(x) —> R(Suggestions)
- F5(x) —> R(Recommendations)
- F6(x) —> R(Reports)
- F6(x) —> R(Graphs)

Algorithm Mapping :

- c1 is worked on by A1 where c1 set of cameras taking snap- shots C
- A1 produces u1 where u1 U
- u1 is worked on by A2,A3,A4 where A1,A2,A3,A4 Algorithms

Problem Type

For the crowd analysis system :

- If a wrong input in the form of no image or image with wrong head count is given , output will be generated.
- If an image with noise is given to process an inaccurate head count,report and suggestion will be generated regardless of the noise
- If correct input is given the correct output will be generated in the form of reports,graphs and suggestions.
- Crowd analysis can be done in polynomial time by processing several images.

Thus it can be inferred that the system is NP Hard and NP hence is NP - Complete.

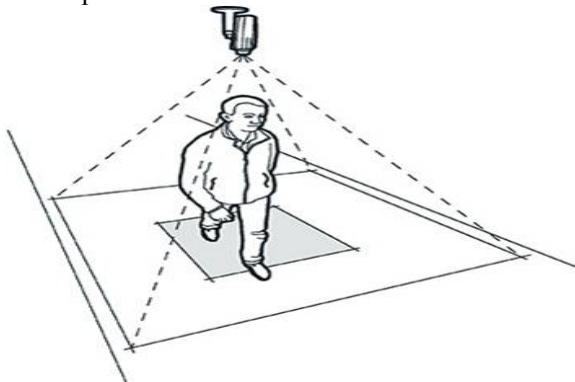


Fig. 4. Possible Implementation

7. SYSTEM ARCHITECTURE

The figure below represents the system architecture of the proposed system.

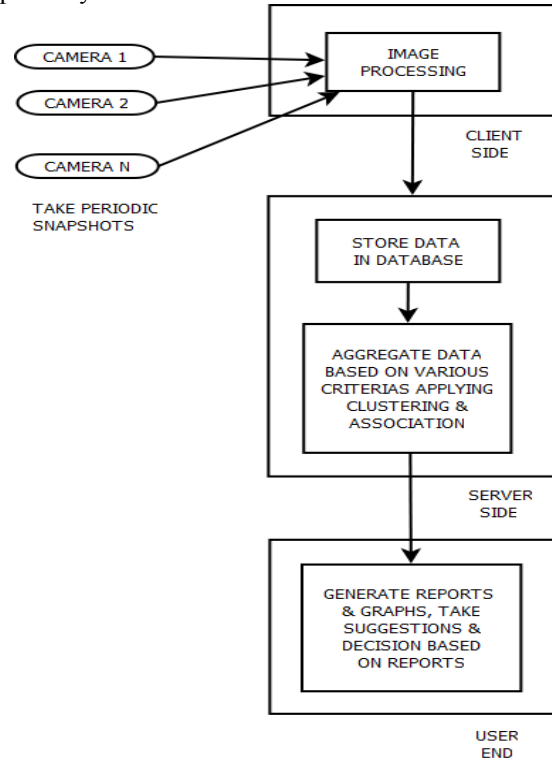


Fig. 5. Proposed System Architecture

8. CONCLUSION

This system will be generic and thus can be used by multiple businesses including shopping malls, supermarkets, hotels and so on. It will be able to rival existing surveillance and monitoring systems by providing more in depth details and information to the concerned personnel involved in the running of the institute.

Owners of the shopping store will be able to make fair business decisions using the automated system compared to the existing system which is prone to human error. Optimized business strategies can be developed to increase business potential and profits.

REFERENCES

- [1] Irshad Ali and Matthew N. Dailey. Multiple human tracking in high-density crowds. Computer Science and Information Management Asian Institute of Technology, 2008.
- [2] Ming Jiang, Jingcheng Huang, Xingqi Wang, Jingfan Tang, and Chunming Wu. An approach for crowd density and crowd size estimation. JOURNAL OF SOFTWARE, 2014.
- [3] Wei Li, Xiaojuan Wui, Raffaele Cappelli, Koichi Matsumoto, and Hua-An Zhao. Crowd density estimation: An improved approach. IEEE, 2010.
- [4] Venkatesh Bala Subburaman, Adrien Descamps, and Cyril Carincotte. Counting people in the crowd using a generic head detector. Image Department, Multitel absI, 7000 Mons, Belgium, 2012.
- [5] Zhe Wang, Hong Liu, Yueliang Qian, and Tao Xu. Crowd density estimation based on local binary pattern co-occurrence matrix. IEEE, 2011.
- [6] Hosub Yoon, Dohyung Kim, Suyoung Chi, and Youngjo Cho. A robust human head detection method for human tracking. IEEE, 2004.
- [7] Zui Zhang, Hatice Gunes, and Massimo Piccardi. An accurate algorithm for head detection based on xyz and hsv hair and skin color models. IEEE, 2008.