A Survey on Multiple Face Detection and Tracking in Crowds

Preeja Priji
M.Tech. Student, Department of CSE, Mohandas College of Engineering and Technology, Anad Thiruvananthapuram Kerala, India

Rashmi S Nair
Asst. Professor Department of CSE, Mohandas College of Engineering and Technology, Anad Thiruvananthapuram Kerala, India

Abstract- One of the main application of image processing is in detection of individuals in a complex uncontrolled scenario. The main difficulties are occlusion, angular pose, different light illumination and partial inclusion. Image processing has become faster and versatile with the development of multicores and faster GPU. The traditional face detection and recognition has come a long way with greater accuracy in different lightings. Application involves detection of criminals/Person of interest from a crowd, presence detection in home automation, audience detection and in robotics. This survey compare and overview some of the main papers presenting different methods of multiple human presence detection in an uncontrolled scenario.

Keywords – Image processing, Multiple Face Detection, Real-time face detection, Tracking.

I. INTRODUCTION

The most important thing is detection of face. We recognize different humans by features of their face. We use image processing to mimic this function. Face detection and face recognition are entirely different tasks. Face detection deals with detection of a face from an image. A face mainly consists of eyes, nose, lips and contour. The edges of a given image are detected and checked for faces and Region of Interest is detected. Face detection tells if there is a face or the region where there is a face in the sample image. Whereas face recognition deals with comparing the content of the ROI with a data set to detect the identity of the individual in the image. This can be done using many algorithms. Genetic or neural.

One of the most important organ in our body for living a social life is our eyes. Eyes are a part of a neural network in brain which transforms the received energy and use it to recognize various patterns and objects such as face, food, disaster, obstacles, etc.

The detection and monitoring of faces and facial features in video files is an important and challenging problem in image processing. The area has many applications in social security, model-based coding, gaze detection, human-computer interaction, teleconferencing, military etc. The main image processing tool we use for image processing is MATLAB or OpenCV.

Different methods use different processing sets and resources. The Face detection is an easy process compared to face recognition, because face recognition depends upon many things such as facial features, environmental lighting, angle of view, etc. The prime use of face detection is in security, presence detection, audience recognition, robotics and crime analysis.
Different techniques are used for different applications.

These applications make use of the most required feature for its optimal performance. Different methods have its own advantages and disadvantages depending on accuracy. As accuracy increases detection speed decreases. The main focus of this survey is on multiple face detection from a number of people.

This survey compares and overview some of the selected work relating to detection multiple faces from a number of individuals in an uncontrolled scenario.

The paper is organized as follow Chapter II overviews 5 papers related to face detection. Chapter III gives an overview comparison of the 5 papers. Chapter IV gives the future scope and modified system that can be implemented using the technologies and Chapter V concludes the survey.

II. COMPARISON OF DIFFERENT FACE DETECTION TECHNIQUES

The papers are explained in ascending order based on the year of publication.

A. Automatic Detection and Tracking of Pedestrians in Videos with Various Crowd Densities

The surveillance camera count in human dense area is increasing at a drastic rate which results in massive amounts of videos to be analysed. Manual observation of crowds and pedestrians in such large data is cumbersome and often impractical which makes automated computer vision methods extremely favourable for this purpose.

Automatic pedestrians tracking is one of the required abilities for computerized analysis. The density of pedestrians highly impacts their status in a video. That is in high density crowds, people mostly occlude each other and usually only some parts of the body of each individual are visible. While the full body or a significant portion of the body is visible in low crowd-density. Hence this require different tracking methods which suite the density of the crowd. The paper [1] present two methods for tracking pedestrians in videos with low and high density of crowds. For videos with low density crowd, [1] first individuals in each video frame is detected using a part-based human detector where occlusion is handled. In the second method, a global data association method based on Generalized Minimum Clique Graphs is used for tracking each.

Paper [1] present two approaches for tracking people in high density of crowds. In First method, scene layout constraint is captured by learning Dynamic Floor Field, Static Floor Field and Boundary Floor Field along with flow of crowd and it is used to track individuals in the crowd. In second method, the tracking is performed utilizing contextual and salient information.

The overview steps involved in [1] is given in Fig 1.

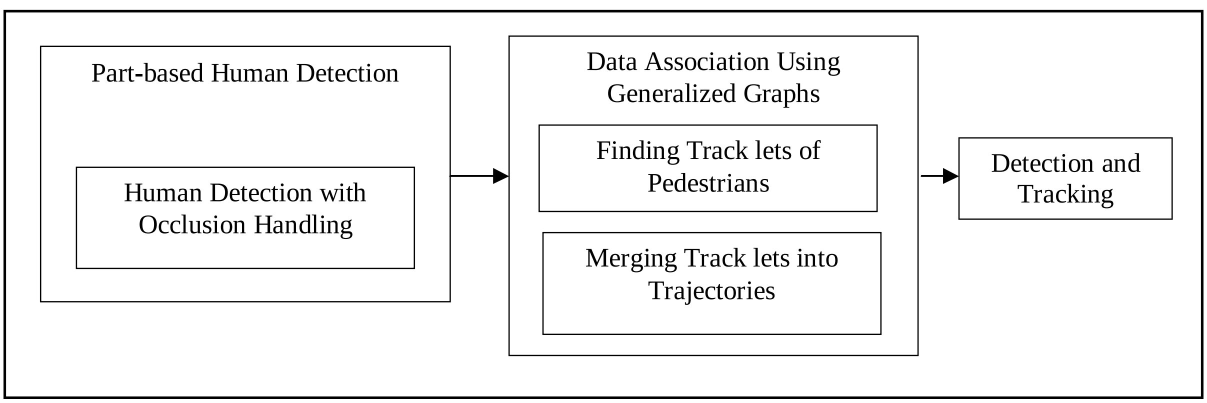

Figure 1 Detection and Tracking of Facial Features in Video Sequences.

B. Crowd Detection from Still Images

This works [2]. focus on detecting and segmenting out crowds of humans from still photos. The goal is to determine if there is a crowd in a sample photo and if so, which portions of the image it includes the crowd. The detection of a crowd form a uncontrolled image environment is useful task in itself. crowd formation can cause delay in underground passages, shopping centers and pedestrian paths, or can an indication of civil unrest. The automotive industry, considers crowds of interest as a potential road hazard. Moreover, crowd segmentation is a useful pre-processing procedure that precedes higher level tasks, such as counting the number of individuals in crowd, and analyzing their behavioral dynamics and interaction. The application where crowd detection can be applied ranges from psychological research and macro-engineering, through to crime prevention and detection.
A crowd can be defined as a group of spatially proximate objects of a certain class. The work specifically considers human crowds, as the type that is usually of most concentrated in practice.

There are many reasons which makes crowd detection challenging. First, limited resolution of images that decreases the possibility of detection. Partial occlusions are prevalent in crowds, and the variation in dress, pose, light makes it difficult. Detection of individuals as the basic building element is not a promising approach [2]. Where as a method that directly looks for multiple people, faces problems of modelling an increased range of variability in their combined appearance, also crowd specific factors such as the spacing of individuals in the crowd that is its density.

Overview steps for work [2] is shown in Fig 2.

Constant scale model generation

↓

Scale-invariant word density

↓

Characteristic response

↓

Detection

Figure 2 Steps involved in [2] Crowd Detection from Still Images

C. *Density-aware person detection and tracking in crowds*

While standard methods such as “scanning-window” methods attempt to localize objects independently, several methods exploit scene context and also relations among objects for improved object detection.

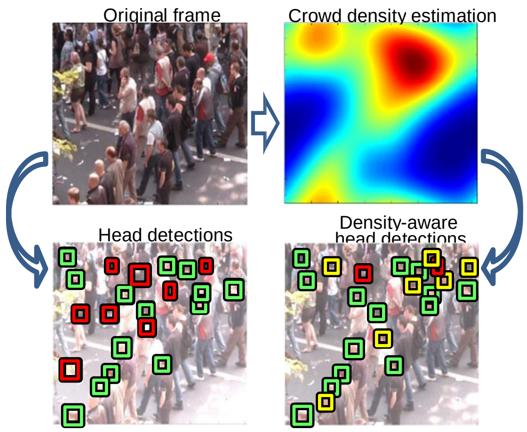

Figure 3. Individual head detections provided by state-of-the-art object detector

The work [3] follows the above-mentioned line of work and extends it to the detection and tracking of people in high-density crowds instead of modelling individual interactions of people, the work uses information at a more global level provided by the geometry of scene and crowd density. Some crowd detection method avoids the hard detection task and attempt to infer person counts directly from low-level image measurements. These methods provide person counts in image regions but is uncertain about the location of detected faces.

The goal and contribution of the analyzed work [3] is to combine these two sources of complementary information for improved person detection and tracking. The prediction behind the method is illustrated in Figure 3 where the constraints of person counts in local image regions helps improvement of the standard head detector.

The method is formulated in an energy minimization framework which combines crowd density estimates with the strength of individual face detections. This energy is minimized by jointly optimizing the density and the location of individual faces in the crowd. The work demonstrates optimization of such leads to significant improvements of state-of-the-art person detection in crowded scenarios with varying densities.

With crowd density cues, the constraints provided by scene geometry and temporal continuity of person tracks in the video is explored and demonstrate further improvements for person tracking in highly crowded scenario. The approach is validated on challenging crowded scenes from multiple video datasets.

The steps are as follows

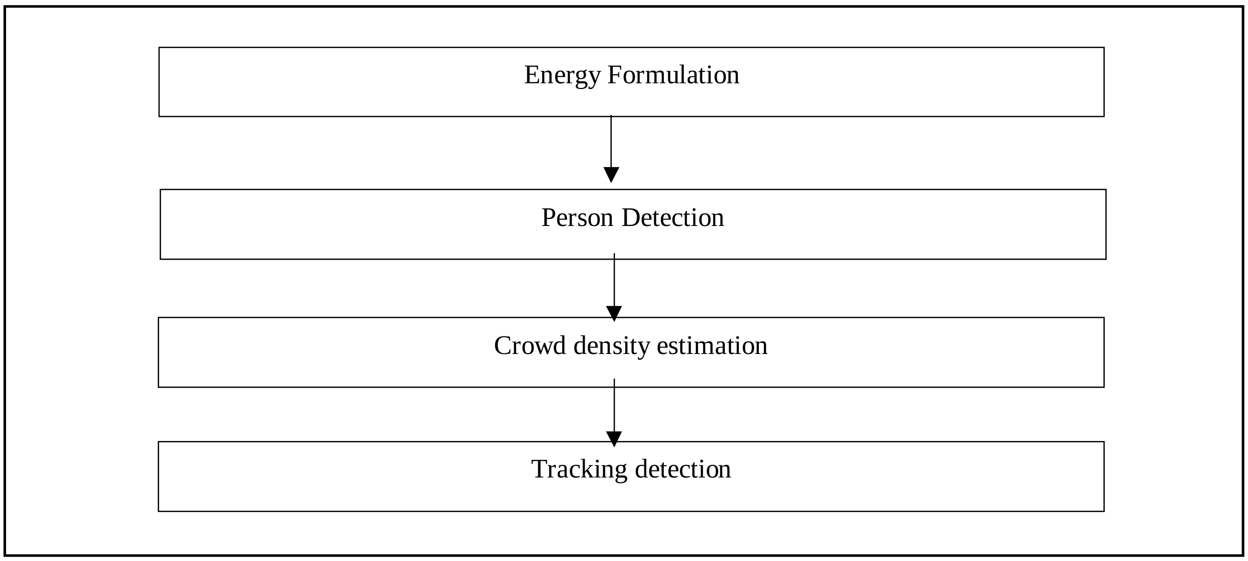

Figure 4 Steps involved in Density-aware person detection and tracking in crowds

D. Unconstrained Face Detection in crowd

The work [4] focus on developing effective features and robust classifiers for unconstrained face detection with arbitrary facial variations. Firstly, a simple pixel-level feature, called the Normalized Pixel Difference (NPD) is proposed. An NPD is the ratio of the difference between any two intensity values of pixel to the sum of the values. The NPD has several desirable properties, such as scale invariance, boundedness, and ability to reproduce the source image. It is easy to compute, involving only one addition, one subtraction, and one division between two values pixels per feature computation.

Secondly a method to construct a single cascade classifier that can effectively deal with complex face contour and handle different pose and occlusions. The weak discriminative ability of NDP is solved by indicating that a subset of NPD features can be optimally selected by Ada Boost learning and combined to create discriminative features in regression tree which is a “divide and conquer” strategy to face and optimized unconstrained face detection in a single classifier, without labelling the views in the training set of the face
images. The proposed face detector is robust to pose variation, occlusion problem, and angular illumination, also to blur and low resolution image.

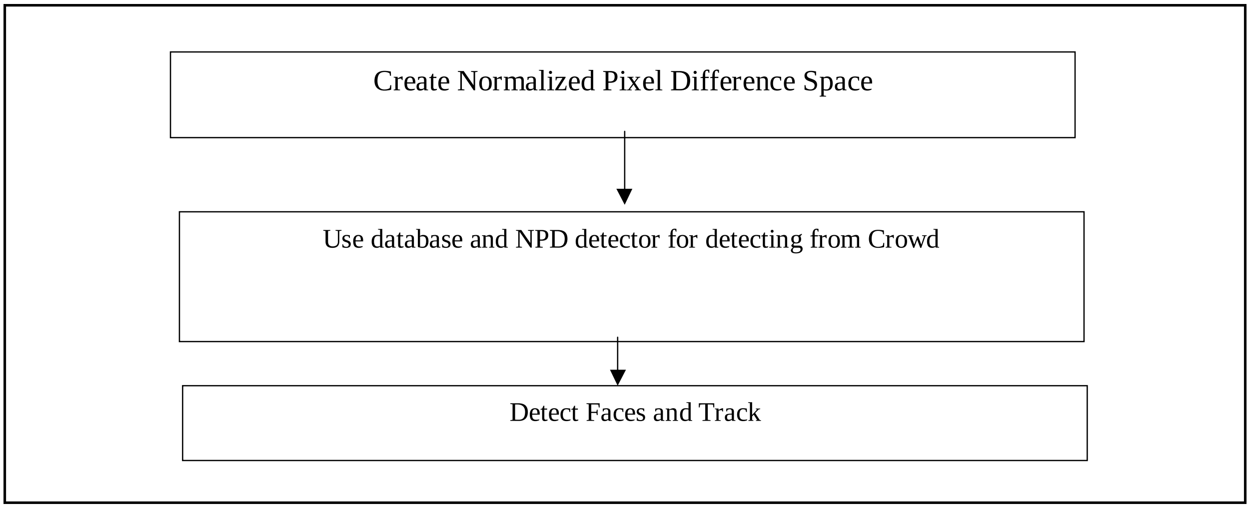

Figure 5 Steps of Unconstrained Face Detection

E. Counting People In The Crowd

This work [5] mainly relies on a head detector to count people from a source image. For detecting the heads from the source image first the point of interest is detected using gradient information from the grey scale image. This approximately locates top portion of the head region to minimize the search region. The points of interest on the source image are masked using a foreground regional space obtained using background subtraction techniques including Vibes and Idiap. Then a sub-window is placed covering the points of interest based on information on perspective calibration and classifies as head or non-head region making use of a classifier. Multiple nearby detections are finally merged to obtain result which is the no of faces.

The overview steps of the proposed system [5] are shown in Fig 6.

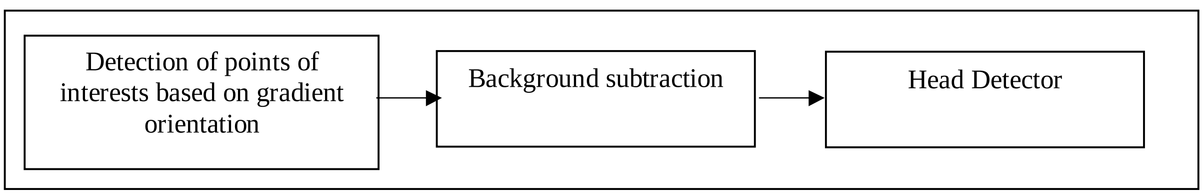

Figure 6. Steps of head Detection

III. FUTURE SCOPES AND WORKS

The different methods use different techniques for detection. The work can be further expanded by recognizing the faces that are detected in real-time. This can be used to recognize criminals, crowd separation, audience detection and for robotics.

IV. CONCLUSION

Image processing is an ever-growing technology with a wide range of application and aids. Multiple face detection is a very complex technique. There are many methods to detect faces. Every one of them have their own advantage and disadvantages. Accuracy decrease with speed and also with image quality. Each method is used for different applications. It is impossible to pin point which method is better. Face detection is a preliminary step in many high-end function like emotion detection, face recognition, lie detection, etc.

REFERENCES

[1] Afshin Dehghan, HaroonIdrees, Amir Roshan Zamir, and Mubarak Shah , “Automatic Detection and Tracking of Pedestrians in Videos with Various Crowd Densities” , Computer Vision Lab, University of Central Florida, Orlando, USA e-mail: adehghan@cs.ucf.edu

[2] U. Weidmann et al. (eds.), Pedestrian and Evacuation Dynamics 2012, DOI 10.1007/978-3-319-02447-9__1,

[3] OgnjenArandjelovi´c, Department of Engineering, University of Cambridge, CB21PZ,UK ,Crowd Detection from Still Images ,InProc. IEEE Conference on Computer Vision and Pattern Recognition(CVPR),2008.

[4] Mikel Rodriguez1,4 Ivan Laptev2,4 Josef Sivic2,4 Jean-Yves Audibert3, Universite Paris-Est ,Density-aware person detection and tracking in crowds, MPEG video compression standard, International Thompson publishing, 2010

[5] Shengcai Liao, Anil K. Jain, Fellow, IEEE and Stan Z. Li, Fellow, IEEE, Unconstrained Face Detection 2013

[6] Venkatesh Bala Subburaman Adrien Descamps Cyril Carincotte , Counting people in the crowd using a generic head detector 2015

[7] C. Wojek, S. Roth, K. Schindler, and B. Schiele, "Monocular 3d scene modeling and inference: Understanding multi-object traffic scenes," in European Conf. Computer Vision, 2010

[8] D. Geronimo, A. M. Lopez, A. D. Sappa, and T. Graf, "Survey on pedestrian detection for advanced driver assistance systems,"IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 32, no. 7, pp. 1239–1258, 2010

[9] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *PAMI*, vol. 34, no. 4, pp. 743–761, 2012.

[10] D. Conte, P. Foggia, G. Percannella, F. Tufano, and M. Vento, "An effective method for counting people in video-surveillance applications." in *VISAPP*, 2011, pp. 67–74.

[11] G. J. Brostow and R. Cipolla. Unsupervised bayesian detection of independent motion in crowds. In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1:594–601, 2006.