Different techniques for Mob Density Evaluation

Tushar kamble[1], Laukik Limbukar[2], Akshay Bhalekar[3], Ameya Wad[4], Prof. Pravin Hole[5]

[1,2,3,4] Dept. of Computer Engineering, Terna Engg. College, Nerul, Maharashtra, India
[5]Professor, Dept. of Computer Engineering, Terna Engg. college, Nerul, Maharashtra, India

---***---

Abstract - *Mob density evaluation means estimation of the overall density of the crowd. Estimating density from crowded images has a wide range of applications such as video. Surveillance, traffic monitoring, public safety and urban planning. Automated Mob density evaluation is an important topic in crowd analysis. The last decades experienced many significant publications in this field and it has been and still a problem for automatic visual surveillance over many years. This paper presents a survey on Mob density evaluation techniques employed for visual surveillance in the perspective of computer vision research. This paper discusses methods such as head detection, Fourier analysis, feature extraction, spatiotemporal detection strategy, and CNN-based Cascaded Multi-task Learning for Mob density Estimation*.

Keywords*: *Density evaluation, Visual Surveillance, Computer Vision, Head detection, Spatio-temporal detection

1. INTRODUCTION

Crowd analysis and monitoring is an essential task in public places to provide secure and safe environments. In past decades, many crowd disasters had taken place due to lack of crowd control management. 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. Crowd density estimation evaluates the capacity of public space design and marks the tendency of abnormal changes of density over the time Human detection in crowd video scenes is getting more proliferation due to the variety of applications in crowd monitoring and tracking [3]. For crowd analysis, automated and semi-automated solutions for density estimation and counting exist in field of computer vision. There is currently a great interest in vision technology for monitoring all types of environments. However, highly dense crowd scenes pose more complications. [3] [1] A huge loss of life and property has been recorded due to stampedes and crowd disasters in the recent years. Many different techniques have been developed for detection and tracking of people in crowded scenes. Images under detection pose interesting challenges to state of the art video analysis techniques for object extraction, basically because of complexity of objects under analysis, and the different effects due to occlusion, shadowing, perspective deformation, complex motion and insufficient image definition. We demonstrate some mob density evaluation techniques. [1] [2] [3] [4]

Some recent incidents [26] [25]:

1. September 29, 2017: At least 22 people were killed and hundreds were injured in a stampede that broke out following heavy rains in a foot over bridge between Parel and Elphinstone road stations in Mumbai, India.

2. September 30, 2017: At least 25 French football fans got injured, four of them seriously, when a barrier collapsed during a match involving home team Amiens SC against Lille OSC at the Stade de la Licorne. The incident occurred around a section housing the away Lille OSC fans.

Fig – 1: Some pictures before crowd crush of Love Parade music event 2010 (Shah et al., 2007).

2. MOB DENSITY EVALUATION TECHNIQUES:

Usually, the problem of evaluation of mob density be segregated into two main approaches: direct and indirect approaches (Conte et al., 2010a).

2.1 Direct Based Approach [1]:

The direct detection-based methods can be further segregated into two approaches: model-based and trajectory-clustering-based approaches. The first approach tries to segment, detect every single individual, and then estimating them using a model or appearance of human shapes. The second approach endeavors to recognize all independent motion in the crowd scene by clustering interest points on people tracked over time and then enumerate the people.

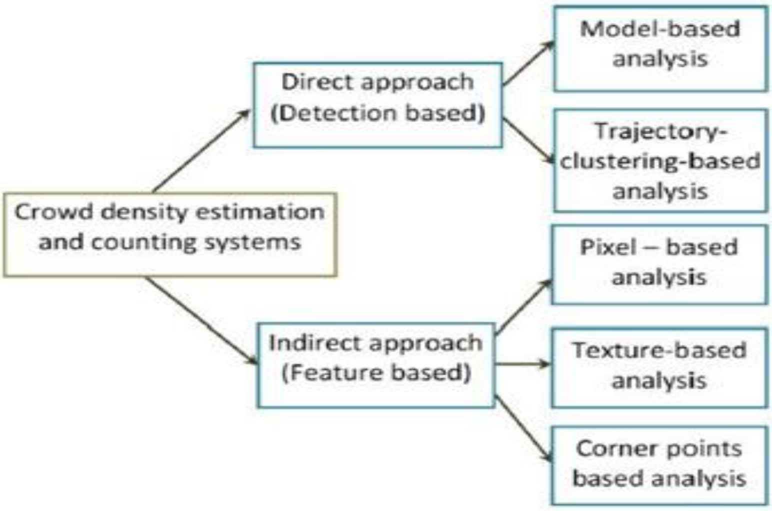
Crowd density estimation and counting systems
Direct approach (Detection based)
Indirect approach (Feature based)
Model-based analysis
Trajectory-clustering-based analysis
Pixel – based analysis
Texture-based analysis
Corner points based analysis

Fig -2: Classification of methods to detect crowd density. [1]

1) Model-based analysis:

Head-like detection: Lin et al. (2001) [1] suggested a recognition approach for the throng (crowd) assessment by wavelet templates and vision-based procedures. In their work, the Haar wavelet transform (HWT) is applied for feature extraction of the head-like contour. Then, this featured area is processed by support vector machines (SVM) to classify it as the contour of a head or not. This method is limited to some complex situation when the contours of the heads are not clear and the computational loading is too heavy, specifically on real-time applications (Lin and Ln, 2006).

2) Trajectory clustering based analysis:

Fig -3: Method proposed by Rabaud and Belongie (2006).

a) *Rabaud and Belongie* (2006) [1] proposed a approach of segmenting motions generated by multiple instances of an person in a crowd. They developeda highly parallelized Kanade Lucas Tomasi (KLT) tracker (Shi and Tomasi, 1994; Tomasi and Kanade, 1991), in order to extract a huge set of low-level characteristic (features) for object commotion detection from the scene. The downside of the method - still based on segmenting individuals on the crowd rather than treating a group as a single entity, as well as it is assumed that the scene is homogeneous.

b) *Cheriyadat et al.* (2008) [1] presented an object detection system rooted on coherent motion region recognition for recognizing and locating individuals in the presence of high density and occlusions. They consider a lone moving object coincide to a lone coherent locomotion region by tracking low-level attributes of objects, and then out-turn a set of independent coherent motion regions as a batch of point tracks. However, in many times, people just remaining static like standing or sitting, exhibiting some occasional articulated movements, which caused false individual detection. In addition, overlapping could take place by sharing more than one individual (person) the same trajectory.

2.2. Indirect detection based [1]:

The indirect methods usually extract many local and holistic characteristics from faction of people in foreground image. These methods are more efficient because detection of features is more effective than the detection of the same (features) of the people.

1) Pixel based [1]:

It is built on the analysis which depends on very local features to compute the number of people in a crowd scene. *Hussain et al.* (2011) [1] propounded an automatic pixel-based crowd density estimation system (CDES) for pictures received from at Masjid-Haram. First, a melding of background removal, using a reference image, and edge detection is applied to frames for feature extraction. Then, this extracted foreground blob pixels are scaled to rectify perspective distortion and act as input for the back propagation neural network (BPNN) to compute the number of people. A supervised training is carried out to segregate the crowd into five distinct batches, from very low to very high. However, false and missed detection cases can be possibly caused by very high crowd density level chiefly because of high occlusion.

2) Texture Analysis [1]:

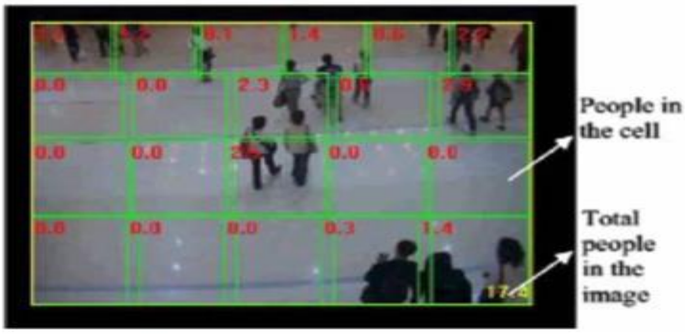

Fig -4: Multi-resolution density cells indicating the estimated numbers of people in individual cell and entire region.

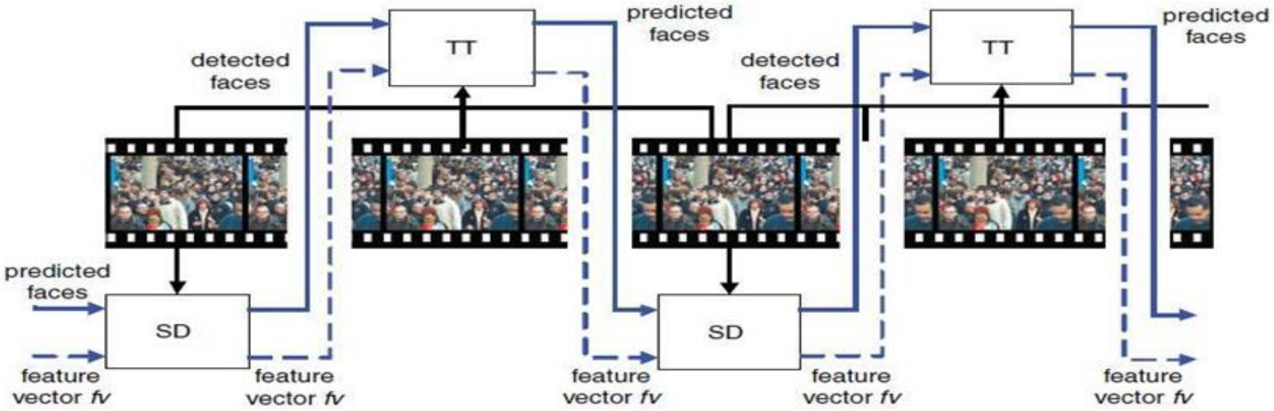

Fig -5: Interaction of SD and TT blocks at processing time.

Texture based approaches explore a more bristly grain and requires the analysis of image patches juxtaposed to pixel-rooted modi operandi

Wu et al. (2006) [1] presented an automatic procedure (approach) to compute locally and globally the crowd density and detect abnormal mob density with the help of texture analysis and support vector machines(SVM).A perspective projection model is used to engender a series of multi-resolution image cells, Figure. Then, the GLDM (Marana et al., 1997) is applied for each cell to deracinate (extract) textural feature vectors. These vectors are rescaled and given to a SVM training system to relate the 15 textural features with the actual density of the scene. However, the drawback of this approach is that when system initial's setup is modified a new training procedure is required.

3) Corner point analysis:

Dittrich et al. (2012) [1] propounded a approach for crowd counting by merging data received from multiple cameras, instead of using single camera, to attenuate the occlusion issue. Their algorithm spots the corrected corner points on the ground plane which are linked with the people present in the scene to estimate their motion vector. They utilized more than one view as a benefit in order to lessen the occurrence rate of occlusions and subsequently the reliability in the enumerating outcome.

3. SOME OTHER APPROACHES

3.1 Mutual feedback scheme:

Mutual feedback scheme for face detection and tracking focused at density estimation [2] in exposition by J.R. Casas, A.Puig Sitjes and P.Puig Folch: It is constructed up of two distinct analysis blocks, the spatial detection (SD) block and the temporal tracking (TT) block, which comprise akin operative structure in terms of their input/output interfaces. A feature vector *fv*, is used to save the parameters of the descries (detected) faces and its advancement with time.

The SD block scrutinizes an original video frame (key frame) utilizing an initial spatial guesstimate of the detected faces and a set of initial characteristics for each face, which are saved in a feature vector. From these statistics (data), the SD block revises the spatial estimate with the help of a spatial segmentation algorithm. The TT block evaluates each original video frame for its motion knowledge until the next key frame. The feature vector is also input given to the TT block. From these details, the TT block revises the estimated regions of support of the detected faces by means of a temporal tracking algorithm.

Mutual Feedback Scheme: The SD block employs temporal statistics to enhance its face detection rates in the absence of increasing the rates of false positives. Temporal information is also used to improve robustness in situations with partial occlusions.

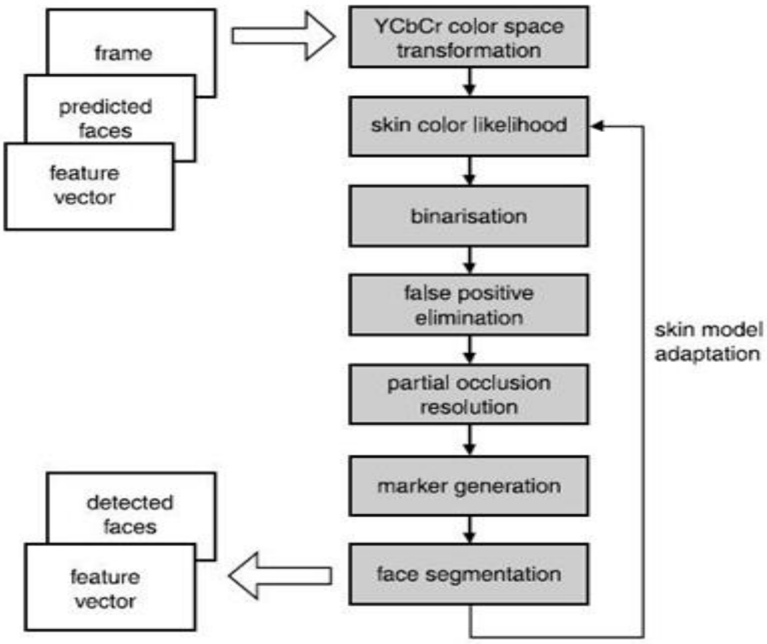

Fig -6: Spatial Detection Algorithm [2]

The TT block supplies information of the number of faces pi that might be included in a given skin region. The TT block enhances its performance because of the analysis computed in the SD block.

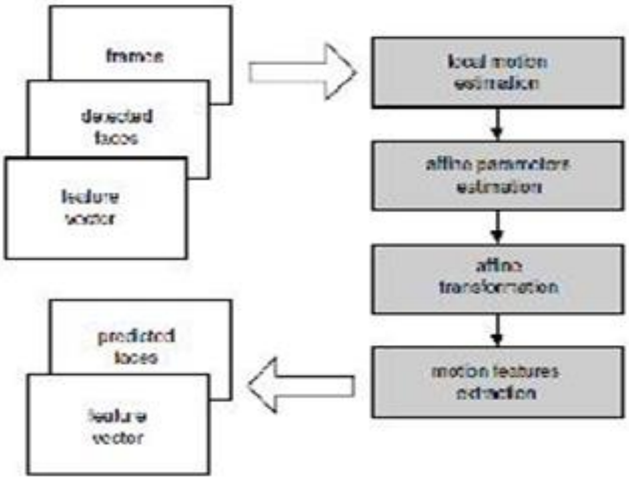

Fig -7 Temporal Tracking Algorithm.[2]

3.2 Human Count Estimation in High Density Crowd Images and Videos [3]:

The major parts of the system are:

a) Head detection: Counts based on human heads appearing in heavily crowded images are calculated by HOG-based feature descriptors.

b) Fourier analysis: crowd is immanently repetitive in nature. Information of an image can be providing by Fourier analysis.

c). Feature extraction: Interest points in crowded images are recognized using SIFT features. The descriptors of such points frame the base strategy that segregates between crowd and non-crowd objects.

d) Fusion we get the ultimate model by fusing the patches from our dataset after getting their counts.

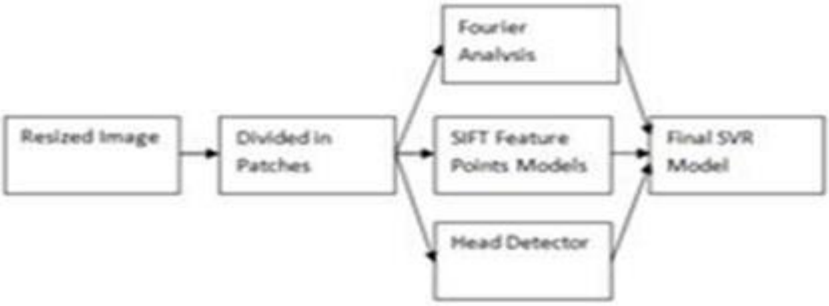

Fig -8 Work-flow of the above system.

3.3 CNN base Cascaded Multi-task learning of High-level Prior and Density Estimation for Crowd-Counting. [4]

This paper reviews a multi-task cascaded CNN network for jointly learning crowd count categorization and density map estimation. By learning to stratify the crowd count into various groups, we are able to incorporate a high level prior into the network which enables it to learn globally relevant discriminative features thereby accounting for large count variations in the dataset. Additionally, the technique employed fractionally stride convolution layers at the end so as to account for the loss of details due to max-pooling layers in the earlier stages thereby allowing us to regress on full resolution density maps. The training of entire cascade is done in an end-to-end fashion.

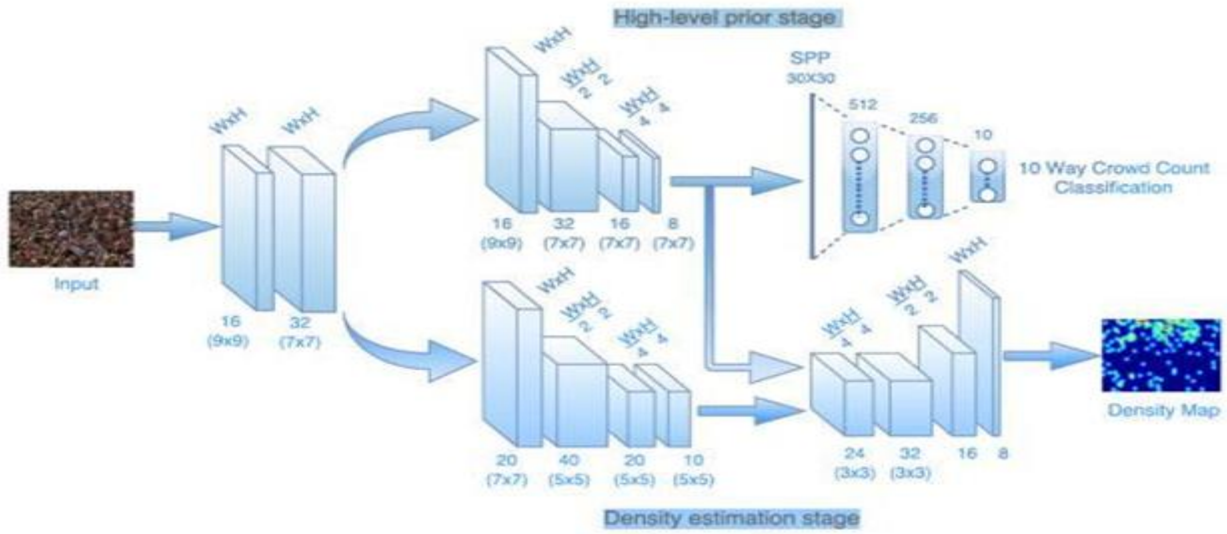

Fig -9: Overview of the proposed cascaded architecture for jointly learning high-level prior and density estimation.

Method	MAE	MSE
Indrees et al.[19]	419.5	541.6
Zhang et al[21]	467.0	498.5
MCNN[22]	377.6	509.1
Onoro et al23[]	465.7	371.8
Walach et al.[24]	364.4	341.4
Proposed Method	322.8	397.9

Table- 1: Comparison results: Estimation errors on the UCF CC 50 dataset.

It can be analyzed from Table 1 that the network achieves the lowest MAE and comparable MSE score.

Some examples of the dataset for crowd density estimation:

1) ShanghaiTech dataset.

The ShanghaiTech dataset was introduced by Zhang et al. [18] and it has 1198 annotated images with a grand total of 330,165 people. The dataset in consideration comprises of two parts: Part A having 482 images and Part B having 716 images.

2) UCF CC 50 dataset

The UCF CC 50 is an extremely challenging dataset introduced by Idrees et al. [19]. The dataset contains 50 annotated images of different resolutions and aspect ratios Crawled from the internet.

3) Pets2009.

The datasets [20] are multisensory sequences having disparate activities of people (crowd).The scenarios are filmed from innumerous cameras and include up to roughly 40 actors.

4. CONCLUSION

In this work, we have presented a review study on Mob density Evaluation methods for surveillance based on computer vision. There are two main different approaches: direct and indirect approaches. In this paper, we also reviewed multi-task cascaded CNN network for jointly learning crowd count classification and density map estimation. A co-operative spatiotemporal procedure aimed at face detection in crowded video sequences can be built on mutual feedback scheme. The method proposed in 'Human Count Estimation in High Density Crowd Images and Videos' Combines information from three different techniques .This paper determine the effective computer based approaches to predict Mob Density based on parameters like accuracy, Execution time and complexity of method.

REFERENCES:

[1] Recent survey on crowd density estimation and counting for visual surveillance .Sami Abdulla Mohsen Saleh, Shahrel Azmin Suandi, Haidi Ibrahim.

[2] Mutual feedback scheme for face detection and tracking aimed at density estimation in demonstrations J.R. Casas, A. Puig Sitjes and P. Puig Folch.

[3] Human Count Estimation in High Density Crowd Images and Videos.Rohit, Vandit Chauhan, Santosh Kumar, Sanjay Kumar Singh.

[4] CNN-based Cascaded Multi-task Learning of High-level Prior and Density Estimation for Crowd Counting, Vishwanath A. Sindagi Vishal M. Patel.

[5] Department of Electrical and Computer Engineering, Rutgers.

[6] Morphological Image Processing: https://www.cs.auckland.ac.nz/courses/compsci773s1c/lec tures/ImageProcessing-html/topic4.htm.

[7] Morphological structuring element : http://in.mathworks.com/help/images/ref/strel.html?reque stedDomain=www.mathworks.com#d119e154052 [8]Morphological skeleton:https://en.wikipedia.org/wiki/Morphological_skele ton

[9]Skeletonization :http://www.inf.u-szeged.hu/~palagyi/skel/skel.html 10)Perspective transformation:https://www.tutorialspoint.com/dip/perspe ctive_transformation.htm

[11]Vanishing point: https://en.wikipedia.org/wiki/Vanishing_point

[12]Watershed:https://en.wikipedia.org/wiki/Watershed_(i mage_processing)

[13]http://in.mathworks.com/help/images/ref/watershed. html?requestedDomain=in.mathworks.com

[14] Recent survey on crowd density estimation and counting for visual surveillance:http://www.academia.edu/28188132/Recent_s urvey_on_crowd_density_estimation_and_counting_for_visual _surveillance

[15]A Survey of Recent Advances in CNN-based Single Image Crowd Counting and Density Estimation.Vishwanath A. Sindagia, Vishal M. Patelb.

[16]Image of crowd (Maratha Morcha) : http://www.dnaindia.com/Mumbai

[17]Picture of Kumbh Mela: http://proof.nationalgeographic.com/2014/02/06/alex-webb-reflecting-on-the-kumbh-mela/

[18]ShanghaiTech dataset: Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma. Singleimage crowd counting via multi-column convolutional neural network.

[19] UCF CC 50 dataset :H. Idrees, I. Saleemi, C. Seibert, and M. Shah. Multi-source multi-scale counting in extremely dense crowd images

[20]Pets2009 dataset: www.cvg.reading.ac.uk/PETS2009/a.html

[22]Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma. Singleimage crowd counting via multi-column convolutional neural network.

[23] D. Onoro-Rubio and R. J. L´opez-Sastre. Towards perspective-free object counting with deep learning. In ECCV, pages 615–629. Springer, 2016.

[24] E.Walach and L.Wolf. Learning to count with cnn boosting. In ECCV, pages 660–676. Springer, 2016

[25]HumanStampedes:https://en.wikipedia.org/wiki/List_of _human_crushes

[26]List of Human stampedes https://en.wikipedia.org/wiki/List_of_human_stampedes.

[27] Crowded Scene Analysis: A Survey Teng Li, Huan Chang, Meng Wang, Bingbing Ni, Richang Hong, and Shuicheng Ya