Crowd Counting, Density Estimation and Analysis of Various Techniques

Ashwin Waradkar, Srivaramangai R

Abstract— Crowd counting and density estimation has many applications such as city planning, pedestrian planning, planning a festival, video surveillance, public safety and designing a smart city. The crowd counting and estimation techniques are also useful in other fields like animal and plant counting (environment survey) in biology for counting microorganisms. The task of crowd counting is difficult because of human nature and the tools used, problems can occur if the dataset used is not correct or does not contain the accurate representation of the crowd. The task of crowd counting and density estimation can be difficult because of the challenges which are perspective, non-uniform density, occlusion, intra-scene and inter-scene variations in scale and weather in spite of that in recent years there are many improvements on the field crowd analysis and giving accurate crowd counting with minimum error. This is due to the availability of datasets which are complex. In this paper, we layout some of the methods used earlier and the current successful methods, we also examine the dataset available for us to use. First, we take a look at the earlier methods used and the benefits of using them. Furthermore, I discuss and determine promising avenues of research in this rapidly evolving field

Index Terms— Crowd Counting, Density Estimation, Genetic Algorithms, Convolutionary Neural Network, Image based Counting.

1 Introduction

Providing accurate estimates of the number of people at a given location, also in real time, is of high importance in the context of city planning and building public transport surveillance and crowd management in the case of natural or man-made disasters. Crowd counting has significant attention from researchers in recent years due to a variety of reasons, such as exponential growth in the world population and the resulting urbanization has led to an increased number of participants in events such as sports, political rallies, public demonstrations or use of public services etc.

Fig.1 Illustration of various crowded scenes and the challenges associated with them. Overlapping of subjects, variation in scale and perspective, clustering can be observed in the images.

- *Ashwin Waradkar is currently pursuing masters degree program in Information Technology, University of Mumbai, India, E-mail: ashwinwaradkar@gmail.com*
- *Srivaramangai R is currently the head of the department of Information Technology, University of Mumbai, India E-mail: rsrimangai@udit.mu.ac.in*

As such it has become crucial to know the accurate count of crowds for better management and security of the public. As any other computer related vision problems there are many difficulties such as image resolution, non-uniform density, perspective etc.(shown in Fig. 1).

Crowd counting focus is to count the number of people during a crowded scene wherever as density estimation aims to map an input crowd image its corresponding density map that indicates the quantity of people per pixel present within the image and therefore the two issues are collectively addressed by researchers. The matter of crowd counting and density estimation is of importance as it can help in improving crowd monitoring and scene understanding. Crowd analysis is a multidisciplinary, crowd counting has various critical applications that are multidisciplinary in nature.

Safety monitoring: As the availability of cameras increases, they can be easily used for surveillance for safety and security purposes in places like tourist spots, temples, festivals, shopping malls, airports, stadiums, political rallies and protests. As such we need an algorithm that can be used in various cases and scenes and perspectives.

Disaster management: In many cases of crowded event like stadium, political rallies and protests are at the risk of crowd related disasters like stampedes or crowd crush can be dangerous and even life threatening. In such scenarios crowd count can be used as a preventive to detect overcrowding and as such measures can be taken [1].

Design of Smart cities: Crowd analysis on public places like bus stops, train stations, shopping malls, airports, gardens etc. can give us an idea where the limitations are present with respect to crowd safety and management. Analyze pedestrian flow at signals at different times of the day and this infor-

mation can be used for optimizing signal-wait times. This can also be used to improve public transport and provide convenience to the public.

Data gathering and analysis: crowd counting technique can be used for collection of data and data interpretation. For example, in a shopping mall counting can be used to analyze public interest as to which store or brand is preferred. It can also be used in store to check popular items and it can be used to optimize queue lengths. The variety of methods and applications has motivated researchers to develop algorithms or tools for crowd counting and density estimation. Additionally, methods developed for crowd counting can be easily extended to counting tasks in other fields such as cell microscopy [6], vehicle counting, environmental survey.

2 Crowd Counting Methods and Techniques

Crowd counting technicques are broadly divided into two major types namely desne and sparse. When there are too much of people at one place covering each and every bit of the space, then it is known as dense crowd whereas if the same set of people are spread over an area then it is sparse crowd. Machine learning techniques have been very extensively used for this on images and videos and most predominantly regression and support vector machine have been used for counting. The type of method used determines the type of crowd counting. In both dense and sparse three major methods are used such as counting by detection, regression and estimating the density.

2.1 Counting by Detection

Human recognition in a crowd by detecting with sensors in images and videos and/or using machine learning algorithms to detect the existence is called counting by detection. This is suitable to sparse crowd rather than dense crowd. Detection is also based on three methods such as full body detection, partial body detection and shape detection. Fig 2 is an example of full body detection.

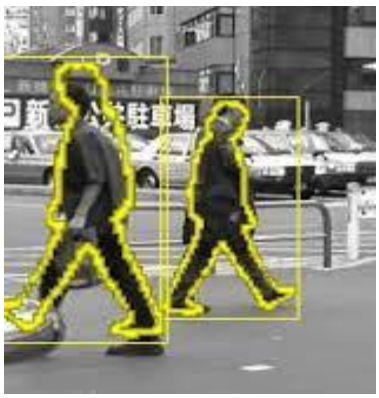

Fig 2. Fullbody detection in sparse crowd

In partial body detection, instead of taking the entire body into consideration, a part of the body like head, neck etc are detected and given to classifier algorithm for processing of recognition. Fig 3 illustrates partial body detection.

Fig 3. Partial Body Detection in dense crowd

The third type is the shape detection. Usually ellipses are drawn arund the shape of the body or part or some feature boundaries and the number is detected using a stochastic process. This is illustrated in Fig 4.

Fig 4. Shape Detection by Drawing Ellipses

2.2 Counting by Regression

When the crowd is dense and too much cluttering in the background of the image or video then instead of counting by detection, it is better to use counting by regression. Low level features of the image are extracted and regression model is applied to detect the count.

2.3 Counting by Estimating the Density

The process of mapping the object density maps to that of local features is called density estimation. In this method, instead of the system learning each and every individual, a group of people's characteristics are learnt. The mapping used can be linear or non linear. Most predominant techniques used in this method are random forest, neural network and decision trees.

3 Survey of Techniques for Crowd Counting

Genetic Algorithms has been used in a wide variety of optimization tasks, including numerical optimization, and combinatorial optimization problems. The algorithm selects a group of individuals in the current population, called parents, who contribute their genes—the entries of their vectors—to their children. Ahmed and et.al [1] have used genetic algorithm along with a simulation model to have an optimized model for evacuation of pedestrian facilties. The application was meant

for crowded exhibition hall where the model will help in need of evacuation, how optimally it could be done. The proposed method has outperformed the existing conventional methods. Three important problems are dealt by Haroon and et.al [2] like counting, density map estimation and localization of people. The research has used optimized CNN algorithms in order to count the crowd and also helping in large number of crowd annotations. Song and et.al [3] have used regression methods to do crowd counting and also spatial temporal analysis of the region. The system developed was able to alert with warning in case of security issues. Muhammed, Sultan and Saleh [4] has proposed a model for counting the number of moving people in a crowded area. They have used Blob analysis and genetic algorithm to count and optimize the different parameters used for counting and density mapping. Tianjun, Qingge and Ning [5] used gradient distribution for selection and classification of groups and was combined with perspective map. The proposed algorithm by the authors gave a better accuracy level of performance in counting the people in crowd when compared with some existing methods. Mark and et.al [6] has propose a CNN based object counter as per the visuals leading to domain specific parameter changes. A very important factor of this research is that a new challenging dataset known as Dublin Cell Counting (DCC) has been created and made available for similar researches. The proposed model works well with the datasets used by the authors such as Shanghaitech and Penguin. Babu Sam and et.al [7] has developed a hierarchical regression tree based CNN model in a recursive pattern where each child created narrows down to the specialities thus performing better than the parent model. A classifier network is trained from the leaf nodes which are specialized children nodes. This classifier network then predicts the counts in that specialized area for the test input images.

Yanyan Fang and et.al [8] have propose in their work a model named Locality-constrained Spatial Transformer Network (LSTN) in order to estimate the density map of the current frame with next frame in the image input. The performance found was better to other methods and was tested with large dataset. End-to-end trainable deep architecture has been used by Liu and et.al [9] where they have combined the features from multiple fields which are of varied sizes. The algorithm was able to perform crowd counting even with different data size. Lingbo Liu and et.al [10] have developed deep recurrent spatial aware network which takes into account the variations in density of the crowd. On applying this technique on Wordexpo and UCF-CC-50 datasets the authors could achieve 12% and 28% accuracy on comparison with the existing techniques referred in the paper. Sindagi, Yasarl and Patel [11] have introduced an unconstrained crowd counting dataset where they used VGG16 as the backbone network. They have proposed Confidence Guided Deep Residual Counting Network (CG-DRCN) which was tested with complex datasets and their own dataset and it has shown minimized errors when compared to previous methods. Loy et al. [14] has done an elaborate comparisons of the techniques used for crowd counting based on video imagery with the same protocol. Tripathi et al. [15] had done a thorough study and reviewed those research work purely based on CNN. The behavioural pattern of the algorithm is studied so as to have new approaches or to improve the methods. Lemptisky et al. [16] in their study of density estimation, have adopted a classifier learning system and did a linear mapping between these density maps and the local features. Zhang and et.al [17] created training set of images by random selection and used density maps of corresponding image patches as the true positive. The total count of the selected training set is calculated by integrating over the density map. Sheng and et.al [18] tried to integrate the semantic information into learning locality-aware feature (LAF) sets for crowd counting. The original pixel space of the image was mapped to the density attribute pixel space feature map and detected the local and spatial information. Zhou and et.al [19] have used a semi-supervised method by utilizing a sub-modular to choose the most appropriate representative frames from the sequences of images to evade redundancy and retain densities. Onoro and Lopez [20] have primarily used a regression function to project the appearance of the image patches onto their corresponding object density maps. The model adopts the same sizes of all patches and the same covariance value of the Gaussian function in the true positive density map generation process, which limits the accuracy when encounters the large scale difference in scenarios.

4. CONCLUSION

In this paper, we have studied the cocnepts of crowd counting and also surveyed the research work already done in this area. Numerous algorithms have been developed and some of them have used existing machine learning algorithms. From the literature survey, it is obvious that CNN is widely used for crowd counting irrespective of dense or sparse crowd. Still there are areas of improving the accuracy of the algorithms.

REFERENCES

[1] Abdelghany, Ahmed, Khaled Abdelghany, Hani Mahmassani, and Wael Alhalabi. "Modeling framework for optimal evacuation of large-scale crowded pedestrian facilities." *European Journal of Operational Research* 237, no. 3 (2014): 1105-1118.

[2] Idrees, Haroon, Muhmmad Tayyab, Kishan Athrey, Dong Zhang, Somaya Al-Maadeed, Nasir Rajpoot, and Mubarak Shah. "Composition loss for counting, density map estimation and localization in dense crowds." In Proceedings of the European Conference on Computer Vision (ECCV), pp. 532-546. 2018.

[3] Song, Hongquan, Xuejun Liu, Xingguo Zhang, and Jiapei Hu. "Real-time monitoring for crowd counting using video surveillance and GIS." In 2012 2nd International Conference on Remote Sensing, Environment and Transportation Engineering, pp. 1-4. IEEE, 2012.

[4] Arif, Muhammad, Sultan Daud, and Saleh Basalamah. "Counting of people in the extremely dense crowd using genetic algorithm and blobs counting." IAES International Journal of Artificial Intelligence 2, no. 2 (2013): 51.

[5] Ma, Tianjun, Qingge Ji, and Ning Li. "Scene invariant crowd counting using multi-scales head detection in video surveillance." IET Image Processing 12, no. 12 (2018): 2258-2263.

[6] Marsden, Mark, Kevin McGuinness, Suzanne Little, Ciara E. Keogh, and Noel E. O'Connor. "People, penguins and petri dishes: adapting object counting models to new visual domains and object types without forgetting." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8070-8079. 2018.

[7] Babu Sam, Deepak, Neeraj N. Sajjan, R. Venkatesh Babu, and Mukundhan Srinivasan. "Divide and grow: Capturing huge diversity in crowd images with incrementally growing cnn." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3618-3626. 2018.

[8] Fang, Yanyan, Biyun Zhan, Wandi Cai, Shenghua Gao, and Bo Hu. "Locality-Constrained Spatial Transformer Network for Video Crowd Counting." In 2019 IEEE International Conference on Multimedia and Expo (ICME), pp. 814-819. IEEE, 2019.

[9] Liu, Weizhe, Mathieu Salzmann, and Pascal Fua. "Context-Aware Crowd Counting." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5099-5108. 2019.

[10] Liu, Lingbo, Hongjun Wang, Guanbin Li, Wanli Ouyang, and Liang Lin. "Crowd counting using deep recurrent spatial-aware network." arXiv preprint arXiv:1807.00601 (2018).

[11] Sindagi, Vishwanath A., and Vishal M. Patel. "Multi-Level Bottom-Top and Top-Bottom Feature Fusion for Crowd Counting." In Proceedings of the IEEE International Conference on Computer Vision, pp. 1002-1012. 2019.

[12] Cheng, Zhi-Qi, Jun-Xiu Li, Qi Dai, Xiao Wu, and Alexander G. Hauptmann. "Learning Spatial Awareness to Improve Crowd Counting." In Proceedings of the IEEE International Conference on Computer Vision, pp. 6152-6161. 2019.

[13] Zhang, Anran, Lei Yue, Jiayi Shen, Fan Zhu, Xiantong Zhen, Xianbin Cao, and Ling Shao. "Attentional Neural Fields for Crowd Counting." In Proceedings of the IEEE International Conference on Computer Vision, pp. 5714-5723. 2019.

[14] C. C. Loy, K. Chen, S. Gong, and T. Xiang, "Crowd counting and profiling: Methodology and evaluation," in MSVAC. Springer, 2013, pp. 347–382

[15] G. Tripathi, K. Singh, and D. K. Vishwakarma, "Convolutional neural networks for crowd behaviour analysis: a survey," The Visual Computer, vol. 35, no. 5, pp. 753–776, 2019

[16] V. Lempitsky and A. Zisserman, "Learning to count objects in images," in NIPS, 2010, pp. 1324–1332.

[17] C. Zhang, H. Li, X. Wang, and X. Yang, "Cross-scene crowd counting via deep convolutional neural networks," in CVPR, 2015, pp. 833–841.

[18] B. Sheng, C. Shen, G. Lin, J. Li, W. Yang, and C. Sun, "Crowd counting via weighted vlad on a dense attribute feature map," TCSVT, vol. 28, no. 8, pp. 1788–1797, 2016

[19] Q. Zhou, J. Zhang, L. Che, H. Shan, and J. Z. Wang, "Crowd counting with limited labeling through submodular frame selection," T-ITS, vol. 20, no. 5, pp. 1728–1738, 2018

[20] D. Onoro-Rubio and R. J. Lopez-Sastre, "Towards perspective-free ´ object counting with deep learning," in ECCV. Springer, 2016, pp. 615–629