An Optimal Approach to Detect the Human Heads using H-MTF in Crowded Scenes

Akanksha Bansal*, Gunjan Verma, Manoj Kumar

Dept. of Computer Engineering and Applications, GLA University, Mathura-India

*Corresponding author

doi: https://doi.org/10.21467/proceedings.114.8

Abstract

Due to the increase in the number of people at crowded places leads to some disaster events, there is a necessity to detect the human heads and estimate the crowd density. The counting of the human heads is quite an immense topic in computer vision and digital image processing. This paper focuses on sample frames that are to be extracted from the crowd video UCF_HDDC and S_HOCK datasets. Our proposed Hybridization-Multiple Target Features (H-MTF) method, detects head objects using three prominent features: texture, color, and shape (T, C, and S). With the help of H-MTF, the optimal value can be estimated to detect the exact spot of the head in a crowded place. By applying two evaluation metrics: (i) Average Precision metric (AvP) and (ii) Average Recall metric (AvR), H-MTF has been compared with the existing methods using 2 different datasets. The results are shown in terms of AvP and AvR and our H-MTF method outcomes best from the existing methods.

Keywords: Crowd, Counting of Head, Density Estimation, Crowd Video.

1 Introduction

In the current era, the crowd is an important research area in computer vision and digital image processing. This research paper is related to the project of "Environmental Monitoring at events". The main objective is to understand the relation between different persons during gatherings at crowded scenes. The "crowd" word refers to a large number of people who lived together in a disorganized and unruly way. In other words, the crowd is a set of people arranged in one group. This whole group is known as mass gatherings. Mass gatherings may include railway stations, shopping malls, buses, religious festivals, public demonstrations, sports events, traffic jams, markets, subways, political rallies, etc. This gathering is too much of the people around one place at the same time.

Fig. 1. Illustrations of crowded scenes (a) Parade (b) Musical concert (c) Sports Stadium (d) Public demonstration.

The formation of the crowd is usually recognized in psychological, social, and tentative actions that must be looking for that [1]. During this gathering, several factors also influence our health. Some problems such as weather, event duration, crowd density, seating arrangements, use of alcohol and drugs, and the location

Proceedings DOI: 10.21467/proceedings.114; Series: AIJR Proceedings; ISSN: 2582-3922; ISBN: 978-81-947843-8-8

of the events such as indoor or outdoor influence illness and rate of injury. These factors will helpful for the injury rates minimum [2]. As shown in Fig. 1. [3] the illustrations of the crowded scene happened mostly in (a) parade, (b) musical concert, (c) sports stadium, and (d) public demonstration of crowded scenes.

From all these factors, we just focused only on crowd density. The crowd density is the total number of people who lived in the given place. But the crowd density is too complex so, we started by counting the number of head objects in crowded scenes [3]. There are two main challenges, occlusion and cluttering.

Detect the human head in crowded places is so difficult task in computer vision. In this paper, the number of human heads are detected and counted in crowded areas such as a parade, sports stadium shown in Fig. 1. Therefore, we are moving with its step-wise process [4]. The important system is the intelligent surveillance system for crowd analysis. The first step focuses on the number of people in the native place that required the low-cost camera hardware, and some algorithms to explore it. There must be a greater number of surveillance cameras installed to improve the situations of crowded places [5]. The camera is the oldest tool [6] and it is too complex to read out all the videos simultaneously. There is an automated system for classifying and monitoring the situation. Many researchers tried to understand various approaches to count the people. To summarizing, the contributions of the paper are:

- In this paper, the proposed Hybridization-Multiple Target Features (H-MTF) method is presented.
- H-MTF detects human head in crowded places.
- H-MTF counts the number of human heads based on specific features: texture, color, and shape (T, C, and S).
- The H-MTF is used to estimate the optimal value.

The following sections are structured as: Section II highlights the motivation along with disaster events in crowded areas. Section III gives the brief challenges for the crowded scene analysis. Section IV provides a detailed review of the related work with its existing techniques including merits and demerits. Section V presents the proposed Hybridization-Multiple Target Features (H-MTF). Section VI discusses the experimental results with evaluation metrics via two benchmark datasets are UCF_HDDC and S_HOCK. Finally, concluding remarks are made in Section VII and lastly, ending with its promising future directions.

2 Motivation

The conventional methods are not appropriate for densely crowded places. The common problems faced are occlusion and cluttering in crowded areas. We observed some of the primary reasons that have fallen on the overcrowding places. Table I. highlights the disaster events that happened due to the mass gatherings. The disaster events lead to a large number of human injuries. Some sample disaster events are discussed in Table I including date and venue, the program event on the mentioned date, the disaster event that happened on that venue, and at last, calamities are given in which the number of human deaths and injuries are mentioned. These disasters happened in many parts of the world [7]. The events that happened in these crowded places resulted in thousands of deaths. In 2004, the most dangerous firework event happened in China which was too bad for us [8]. These events mainly discussed the incidents which motives to study the counting of crowd funding. These incidents studied the basic disaster event including with its deaths and injuries.

TABLE I. SUMMARIZATION OF DISASTER INCIDENTS.

Date & Venue	Program Event	Disaster Event	Calamities
11th May 1985 Ibrox Stadium, Bradford, UK	Football Match	Barriers overcrowding.	Deaths:66 Injured:766
3rd July 1990 Mecca, Saudi Arabia	The Haj	Overcrowding and rush towards exit resulting in a stampede.	Deaths:1400 Injured:1937
5th May 1992 Bastia, Corsica	Football Match	Overcrowding of temporary stand which resulted in the collapse	Deaths:17 Injured:1900
28th October 1998 Gothenburg, Sweden	Night Club Event	Overcrowding at the venue. Stairway fire at an emergency exit.	Deaths:63 Injured:213
11th Feb 2004 Miyun District, China	Firework Event	Overcrowding has resulted in a stampede.	Deaths:37 Injured:24

3 Challenges

Due to the overcrowding in public places, the persons come in trouble to handle the situations. These problems occur whenever an object creates disturbances or hurdles in the respective areas [9]. These hurdles such as trees, light poles, etc., and it is very hard to resolve out. These complex parts solved when all the objects didn't disturb each other and solve the problem simultaneously [10-11]. There are two main challenges for crowded areas: a) occlusion and b) cluttering that are described below in detail. Fig. 2. [9] shown the challenges of crowd analysis in crowded places.

- Occlusion: The term occlusion occurs when there is an occurrence of a disturbance in which we cannot maintain the distance between two objects. But, somehow, when there are no barriers then this challenge will be successful.
- Cluttering: The term cluttering refers to the disarrangement between things. The clutter means disordering the object to cover up in the right way. Somehow, this challenge is complicated but it is simple when the objects are not too far from each other.

Fig. 2. Challenges at crowded places.

Proceedings DOI: 10.21467/proceedings.114
Series: AIJR Proceedings
ISBN: 978-81-947843-8-8
ISSN: 2582-3922

4 Related Work

This section reviews existing renowned methods to detect the head. Most techniques of the crowd analysis such as density estimation, pedestrian walkway, crowd recognition, etc. Many approaches have been implemented to monitor the surveillance cameras in crowded places [12]. The overall summary of the existing renowned methods with their merits and demerits listed in Table II. The author Zhao et al. [13] proposed an algorithm to model the human shape and describes the shape with its color histogram feature. The author used the Gaussian distribution model to manipulate their work to enhance the segmentation scheme. Brostow et al. [14] discussed the probabilistic framework. The method was compared to the existing approaches to presents the tracking for detection in crowd dynamics. Zhan et al. [15] presented a survey paper in 2008 on crowd analysis methods in computer vision systems.

TABLE II. SUMMARY OF EXISTING WORK BASED ON VARIOUS TECHNIQUES.

Ref/ Year	**Author**	**Approach**	**Merits**	**Demerits**
2014 [16]	Haroonld ress et al.	Motion Occurrence	Handling the dynamic and dense crowd flow.	It used only spatial information.
2015 [17]	Mukherje e et al.	Lagrangian	Handle fast and dense crowd flow.	Heavy occlusions are created so that there is a cumulative effect of the motion.
2011 [18]	Hajer-Fradi et al.	Density map based	Tracking is too fast because surveillance cameras are installed.	The people are not comfortable in the zone due to the environmental changes.
2012 [19]	Irshad and Mathew	Classification	Track multiple humans in densely crowded areas.	The approach is capricious.
2015 [20]	Cao et al.	Convolutional Neural Network	The tracking metrics are more robust.	Requires a large amount of data and too much uniformity.

5 The Proposed Method

In this section, we discuss the details of the proposed Hybridization-Multiple Target Features (H-MTF).

5.1 Hybridization Based on Multiple Target Features (H-MTF)

This section presents the proposed H-MTF. H-MTF consists of four phases [7]. These phases are 1) input crowded video, 2) reference point extraction, 3) an optimal head finding using H-MTF, 4) to calculate an optimal value by taking three hybrid parameters that are, texture, color, and shape (T, C, and S) [21]. The H-MTF is just shown in Fig. 3. [7]

- Phase 1: In phase 1, focusing on sample frames that are to be collected from the video dataset.
- Phase 2: In phase 2, the reference points are to be extracted from the corresponding video frames.
- Phase 3: In phase 3, focusing on finding an optimal value using H-MTF.
- Phase 4: In phase 4, T, C, and S parameters are used to evaluate an optimal value.

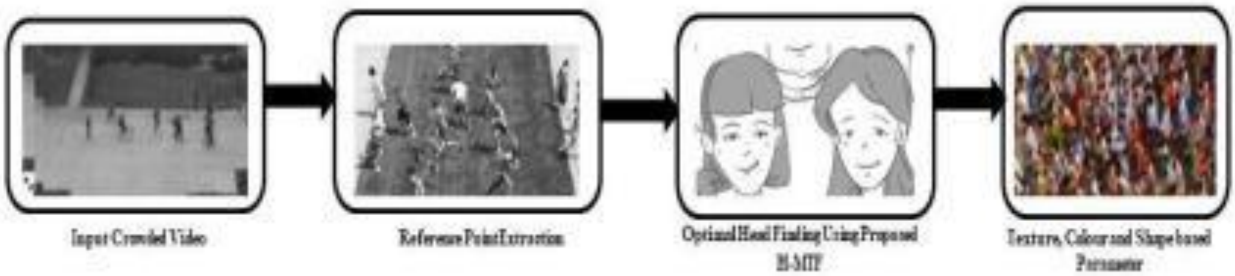

Fig. 3. The Proposed Hybridization-Multiple Target Features (H-MTF).

5.2 Extraction of Video Frames and Evaluation of Reference Point

Firstly, the input crowded video, V is read out and then frames are extracted. After that, each frame of the input video, $V \in (v_i;\ 1 \leq i \leq Z)$ is given as input of the H-MTF. For instance, an input video contains the number of frames, Z. Firstly, we had taken the reference points, Q_k that are chosen randomly. With the help of Q_k first frame is obtained, v_1. This v_1 is used to find the exact head part of the human object that is presented in the first frame [7]. Q_k is computed using equation (1):

$$Q_k = (x_k, y_k)\ ;\ 1 \leq x_k \leq M\ ,\ 1 \leq y_k \leq N \qquad (1)$$

where $M \times N$ is the size of the image.

(x_k, y_k) are the spatial coordinates.

5.3 H-MTF

H-MTF is used for counting the head part in crowded scenes. We considered the unique features of computer vision methods. These features will be beneficial for estimating an optimal value [22-23].

- The first feature is texture: This feature is utilized in the H-MTF for analyzing the head object of a human. The texture word refers to the patterns of the objects. In this, we will more clearly the similarities and differences of the head and then locate the head part of human objects. The texture feature is represented by Tj .
- The second feature is color: This feature is utilized in the color histogram to differentiate the head part. In this, the color feature varied from person-to-person. The color of the hair is changed from brown to black or black to brown. The color feature is represented by Cj.
- The third feature is shape: This feature is utilized to recognize the outline of the objects. The shape feature of any head object is a very powerful native feature that is used to categorize the shape of the head. The shape feature is represented by Sj.

Using equation (2) we can compute the value of $O(Q_k)$ by three parameters as T_j, C_j, and S_j.

$$O(Q_k) = a.T_j + \beta.C_j + \gamma.S_j \qquad (2)$$

where a, β, γ are weighted constants.

6 Results and Discussions

This section gives an overview of the experimental setup of the proposed H-MTF. The publicly available datasets namely UCF_HDDC [24] and S_HOCK [26] are used to detect the human head given by two evaluation metrics- AvP, AvR [7]. The results are compared from the existing method with our proposed H-MTF. Table III and Table IV are based on evaluation metrics given in sub-section C.

Proceedings DOI: 10.21467/proceedings.114
Series: AIJR Proceedings
ISBN: 978-81-947843-8-8
ISSN: 2582-3922

6.1 Experimental Setup

The proposed H-MTF implementation settings are presented in detail in this section. H-MTF is implemented and trained on GPU with system configurations of Intel Xeon Gold 5222 3.8 GHz Processor, Dual Nvidia Quadro RTX4000 8 GB Graphics, Windows 10 Pro Operating System.

6.2 Dataset Description

This section gives a brief description of the high-density crowd video datasets available online on the websites. Fig. 4., [24,26] shows the sample frames of the (a) UCF_HDDC and (b) S_HOCK dataset. The UCF_HDDC dataset is taken from "University of Central Florida: Human Detection in Dense Crowds" which can be downloaded from https://www.crcv.ucf.edu/research/data-sets/ [24]. The S_HOCK dataset is taken from "Spectators Hockey" which can be downloaded from http://vips.sci.univr.it/dataset/shock/ [26].

- UCF_HDDC: Most of the sample images are gathered from marathon ground. This is one of the largest and challenging dataset which is used for human detection in high-density crowds. The sample frames are captured from high-rise buildings so that the images are too blurred with a low-resolution camera [25]. Similarly, the occlusion and lighting conditions also varied due to the camera resolution. It means that these challenges will affect the images in the UCF_HDDC dataset. So, that, the number of pixels in the images is a fewer loss. The sequence of sample images contains 108 frames.

- S_HOCK: Besides, the sample images are gathered from all over the world. This is one of the huge dataset which is used for human detection in high-density crowds. The sample frames are captured from four different match tournaments via five cameras [26]. The S_HOCK dataset was held in Italy (Trento) during the 26th Winter Universiade. Some of the instances such as faults, shots on goals, etc. It means that the S_HOCK dataset will affect the images. So, that, the number of pixels in the images is a fewer loss. The sequence of sample images contains 930 frames. Meanwhile, the 75 video sequences are covered within these frames. Each sequence frame is covered in 31 seconds.

Fig. 4. Sample frames of (a) UCF-HDDC and (b) S_HOCK dataset.

6.3 Evaluation Metrics

Two evaluation metrics are used to evaluate the system: AvP and AvR.

AvP: The precision metric gives the closeness value of the target feature. Precision computes quality. The higher is the precision value, the documents are to be more relevant. The chances of the relevant are more than the irrelevant. AvP is the average precision of data and this data is not precise so much. AvP is also known as Mean Average Precision (MAP) and accuracy.

$AvP = TP / TP + FP$

where TP is True-Positives.

FP is False-Positives.

AvR: The recall metric is also known as sensitivity and Mean Average Recall (MAR). Recall computes quantity. The higher is the recall value, the more documents are to be most relevant. The chances of the relevant are most than the irrelevant. The recall is the fraction of the relevant documents that are to be retrieved.

$AvR = TP / TP + FN$

where *TP* is True-Positives.

FN is False-Negatives.

6.4 Qualitative & Quantitative Evaluation

6.4.1 Qualitative Evaluation:

Fig. 5. represents the detecting results of (a) UCF_HDDC and (b) S_HOCK dataset. It can be observed that the proposed H-MTF successfully detects the human heads in crowed video. The dataset used to obtain the qualitative evaluation [24] and [26]. In Fig. 5. red color represents the detection of human heads in crowded scene. This proves that the proposed H-MTF method is able to detect human heads based on three prominent properties.

Fig. 5. Visual results of (a) UCF-HDDC and (b) S_HOCK dataset.

6.4.2 Quantitative Results:

The quantitative results of the proposed H-MTF method are discussed in this section by comparing them with the existing three existing methods which are shown in Table III and Table IV. Table III and Table IV, reported the performance of different methods using two evaluation metrics as Average Precision (AvP), Average Recall (AvR).

Table III compares the proposed H-MTF method with the existing three methods. These methods are CN-HOG [26], HOG-LBP [27] and DPM [28] evaluated using 3 sample images of the UCF_HDDC dataset. The method CN-HOG [26] added two methods likewise, Color Name descriptor as a feature and Histogram of Gradient feature used for detecting the object [26, 28]. The method HOG-LBP [27] computes HOG and LBP features for human head detection. The method DPM [28] stands Deformable Part Model used for 3-D poses of the human to detect the head objects.

Table IV shows the comparison of the proposed H-MTF method with the renowned existing three methods. These methods are DPM [28], ACF [29] and CUBD [30] evaluated using 3 sample images of the S_HOCK dataset. The method ACF [29] stands Aggregate Channel Features used for color gradients. The method CUBD [30] stands Calvin Upper Body Detector and it is the concatenation of DPM and Viola-

Jones face detector. The outcome of the proposed H-MTF gives the best performance in comparison to CN-HOG, HOG-LBP, DPM, ACF, CUBD methods. The AvP and AvR [31] value of the H-MTF gives better relevant results from the existing ones [32-35].

TABLE III. PERFORMANCE OF DIFFERENT METHODS USING UCF_HDDC [24] DATASET.

Dataset Name	**Methods**	**Average Precision (AvP)**	**Average Recall (AvR)**
UCF_HDDC	CN-HOG [26]	0.35	0.28
	HOG-LBP [27]	0.16	0.21
	DPM [28]	0.33	0.41
	Proposed H-MTF	0.62	0.59

TABLE IV. PERFORMANCE OF DIFFERENT METHODS USING S_HOCK [26] DATASET.

Dataset Name	**Methods**	**Average Precision (AvP)**	**Average Recall (AvR)**
S_HOCK	DPM [28]	0.52	0.44
	ACF[29]	0.50	0.63
	CUBD [30]	0.85	0.32
	Proposed H-MTF	0.89	0.74

7 Conclusion

This paper proposed H-MTF, an objective function using 3 prominent properties i.e., texture, color, and shape (T, C, and S) to compute an optimal value. The proposed H-MTF detects the human heads. Moreover, in this paper, frames are extracted from the UCF_HDDC and S_HOCK datasets with the help of reference points. These frames further help in head detection. In the future, the work can be extended by using the same objective function and track people's behavior in crowded places. Moreover, the results of the proposed H-MTF are compared with the existing 3 methods. It can be observed that our proposed H-MTF gives better performance from the existing methods. And our results are more promising.

References

[1] Sindagi, Vishwanath A., and Vishal M. Patel. "A survey of recent advances in cnn-based single image crowd counting and density estimation." *Pattern Recognition Letters* 107 (2018): 3-16.

[2] Pandey, Anurag, Mayank Pandey, Navjot Singh, and Abha Trivedi. "KUMBH MELA: A case study for dense crowd counting and modeling." *Multimedia Tools and Applications* 79, no. 25 (2020): 17837-17858.

[3] Li, Bin, Jian Zhang, Zheng Zhang, and Yong Xu. "A people counting method based on head detection and tracking." In *2014 International Conference on Smart Computing*, pp. 136-141. IEEE, 2014.

[4] Nguyen, Vy, and Thanh Duc Ngo. "Single-image crowd counting: a comparative survey on deep learning-based approaches." *International Journal of Multimedia Information Retrieval* 9, no. 2 (2020): 63-80.

[5] Khan, Sultan Daud, and Saleh Basalamah. "Sparse to Dense Scale Prediction for Crowd Couting in High Density Crowds." *Arabian Journal for Science and Engineering* (2020): 1-15.

[6] Ranjan, Viresh, Hieu Le, and Minh Hoai. "Iterative crowd counting." In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 270-285. 2018.

[7] Kumar, M., and C. Bhatnagar. "Zero-stopping constraint-based hybrid tracking model for dynamic and high-dense crowd videos." *The Imaging Science Journal* 65, no. 2 (2017): 75-86.

[8] Kumar, Manoj, and Charul Bhatnagar. "Crowd behavior recognition using hybrid tracking model and genetic algorithm enabled neural network." *International Journal of Computational Intelligence Systems* 10, no. 1 (2017): 234-246.

[9] Singh, Upasana, and Manoj Kumar Choubey. "Motion Pattern Recognition from Crowded Video." In *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*, pp. 431-435. IEEE, 2020.

[10] Liu, Weizhe, Mathieu Salzmann, and Pascal Fua. "Context-aware crowd counting." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5099-5108. 2019.

[11] Cao, Xinkun, Zhipeng Wang, Yanyun Zhao, and Fei Su. "Scale aggregation network for accurate and efficient crowd counting." In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 734-750. 2018.

[12] Vora, Aditya, and Vinay Chilaka. "FCHD: Fast and accurate head detection in crowded scenes." *arXiv preprint arXiv:1809.08766* (2018).

[13] Zhao, Tao, and Ramakant Nevatia. "Bayesian human segmentation in crowded situations." In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, vol. 2, pp. II-459. IEEE, 2003.

[14] Brostow, Gabriel J., and Roberto Cipolla. "Unsupervised bayesian detection of independent motion in crowds." In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 1, pp. 594-601. IEEE, 2006.

[15] B. Zhan, D. N. Monekosso, P. Remagnino, S. A. Velastin, and L.-Q. Xu, “Crowd analysis: A survey,” *Mach. Vis. Appl.*, vol. 19, nos. 5–6, pp. 345–357, 2008.

[16] Idrees, Haroon, Nolan Warner, and Mubarak Shah. "Tracking in dense crowds using prominence and neighborhood motion concurrence." *Image and Vision Computing* 32, no. 1 (2014): 14-26.

[17] Mukherjee, Sudipto, Debdipta Goswami, and Sarthak Chatterjee. "A Lagrangian approach to modeling and analysis of a crowd dynamics." *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 45, no. 6 (2015): 865-876.

[18] Rodriguez, Mikel, Ivan Laptev, Josef Sivic, and Jean-Yves Audibert. "Density-aware person detection and tracking in crowds." In *2011 International Conference on Computer Vision*, pp. 2423-2430. IEEE, 2011.

[19] Ali, Irshad, and Matthew N. Dailey. "Multiple human tracking in high-density crowds." *Image and vision computing* 30, no. 12 (2012): 966-977.

[20] Cao, Lijun, Xu Zhang, Weiqiang Ren, and Kaiqi Huang. "Large scale crowd analysis based on convolutional neural network." *Pattern Recognition* 48, no. 10 (2015): 3016-3024.

[21] 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.

[22] Swathi, H. Y., G. Shivakumar, and H. S. Mohana. "Crowd behavior analysis: A survey." In *2017 international conference on recent advances in electronics and communication technology (ICRAECT)*, pp. 169-178. IEEE, 2017.

[23] Kalel, D. S., P. M. Pisal, and R. P. Bagawade. "Color Shape and Texture feature extraction for Content Based Image Retrieval System: A Study." *International Journal of Advanced Research in Computer and Communication Engineering,* 5, no. 4 (2016): 304.

[24] Lamba, Sonu, and Neeta Nain. "Crowd monitoring and classification: a survey." In *Advances in Computer and Computational Sciences*, pp. 21-31. Springer, Singapore, 2017.

[25] Li, Ai Min, Pil Seong Park, and Ye Hong Chen. "An adaptive particle filter tracking method based on homography and common FOV." In *Proceedings of the 2012 ACM Research in Applied Computation Symposium*, pp. 126-131. 2012.

[26] https://www.crcv.ucf.edu/research/data-sets/

[27] Wang, Xiaoyu, Tony X. Han, and Shuicheng Yan. "An HOG-LBP human detector with partial occlusion handling." In *2009 IEEE 12th international conference on computer vision*, pp. 32-39. IEEE, 2009.

[28] http://vips.sci.univr.it/dataset/shock/

[29] Dollár, Piotr, Ron Appel, Serge Belongie, and Pietro Perona. "Fast feature pyramids for object detection." *IEEE transactions on pattern analysis and machine intelligence* 36, no. 8 (2014): 1532-1545.

[30] Khan, Sultan Daud, Habib Ullah, Mohib Ullah, Nicola Conci, Faouzi Alaya Cheikh, and Azeddine Beghdadi. "Person head detection based deep model for people counting in sports videos." In *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1-8. IEEE, 2019.

[31] Pedersoli, Marco, Andrea Vedaldi, Jordi Gonzalez, and Xavier Roca. "A coarse-to-fine approach for fast deformable object detection." *Pattern Recognition* 48, no. 5 (2015): 1844-1853.

[32] Felzenszwalb, Pedro F., Ross B. Girshick, David McAllester, and Deva Ramanan. "Object detection with discriminatively trained part-based models." *IEEE transactions on pattern analysis and machine intelligence* 32, no. 9 (2009): 1627-1645.

[33] Khan, Fahad Shahbaz, Rao Muhammad Anwer, Joost Van De Weijer, Andrew D. Bagdanov, Maria Vanrell, and Antonio M. Lopez. "Color attributes for object detection." In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3306-3313. IEEE, 2012.

[34] Ali, Saad, and Mubarak Shah. "Floor fields for tracking in high density crowd scenes." In *European conference on computer vision*, pp. 1-14. Springer, Berlin, Heidelberg, 2008.

[35] Sindagi, Vishwanath A., and Vishal M. Patel. "Multi-level bottom-top and top-bottom feature fusion for crowd counting." In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1002-1012. 2019.*Automated segmentation of leukocytes using marker-based watershed algorithm from blood smear images*