How to Cite:
Sohani, M. G., & Patekar, S. A. (2022). An improved crowd detection system using text and image filtration methods for precautions from COVID-19. *International Journal of Health Sciences*, *6*(S1), 2046–2056. https://doi.org/10.53730/ijhs.v6nS1.4955

An Improved Crowd Detection System Using text and Image Filtration Methods for Precautions from COVID-19

Mandar Ganesh Sohani
PhD Scholar Department of Computer Engineering, Vidyalankar Institute of Technology University of Mumbai, India

S. A. Patekar
Guide & Professor Department of Computer Engineering, Vidyalankar Institute of Technology University of Mumbai, India

***Abstract*---**The text feature is an important descriptive feature of many image analysis applications. Objectives of this study aim to determine the different texture characteristics of them estimation and calculation of population density. In this paper, we have various reviews of the texture and their own have been extensively reviewed, a different combination that is possible to test their effectiveness in crowds of pedestrians. Divide into two categories and retreat. A framework has been proposed to evaluate performance of all aspects of the density coefficient of human density as well to count. According to the framework, the input images are categorized into blocks and blocks into cells of different sizes, having various levels of overcrowding. Because of a distorted view of people, people's visibility near the camera contributes greatly to this feature vector than distant people. Therefore, the features released are usually using a standard visual map of the scene. In the first stage, picture blocks are classified using multiple classes SVM has been at a different level of congestion. In the second phase Gaussian Process Deactivation is used to restore undoing low-level features for calculation. Various texture features and their possible combinations checked on publicly available dataset.

***Keywords*---**human detection, shadow elimination, partial blocking handling, color correlogram, histogram of oriented gradients (HOG).

Introduction

Increased use of computer technology has encouraged human discovery as an active research field. Personal detection in a video surveillance system has

Corresponding author: *Sohani, M.G.*
Manuscript submitted: 27 Nov 2021, Manuscript revised: 09 Feb 2022, Accepted for publication: 18 March 2022

major application features that include human appearance, patient fall detection and a smart user interface (wiimote, kinect, smart TV). The discovery of a person is a profound and difficult issue because of two challenges: 1) The diversity of categories of people such as appearance, dress, skin color and posture; 2) External problems such as uneven light and dense background. Current recruitment frameworks can be divided into two processes. One process uses a sliding window, while the other process uses partial-based detection.

A window-based sliding process can be upgraded in two areas: Designing additional visual features to improve level of acquisition and use of effective training methods to learn advanced designers. The most commonly used features include the Haar wavelet, HOG [11], shapelet, histogram of edge orientation (EOH), edgelet, regional covariance [6] and LBP.

In [3], a multidisciplinary transformation feature is proposed that incorporates various regional features of the LGP, LBP and HOG. The proposed feature indicates the intensity of local light changes. However, the high size of the hybrid feature increases the complexity of the computer. A polar-based shape feature is generated and used by SVM in the classification [4]. However, system detection is limited to the upper part of the human body. The features of a prominent object are photographed by combining the different thickness of all pixels with the characteristics associated with the texture [1]. Although these multidimensional features capture a large degree of information about an object, the proposed method determines certain key thresholds based on the theory. What has made the road weak when dealing with outdoor issues such as light changes and background clutter. Several sections have been approached for public comment. Many efficient detection detectors often use different variables to strengthen algorithms [7], different types of SVMs or Neural networks. In contrast to body-embracing frameworks, organ-based acquisitions [4], [5] are better suited to manage partial closure. However, the efficiency of improved acquisitions also increases accounting costs. For example, the multi-kernel reading framework (MKL) presented in [9] takes approximately 67 seconds to process each framework. Background removal [8], [10] has been a long-term research field. However, the efficiency of these processes decreases significantly when the ROI does not move for a long time.

This paper proposes a framework to find people trapped by front extraction from the back using the back removal process. The main emphasis of this paper is to remove the shadow regions from the front in order to obtain an accurate ROI. Shadows can be defined as part of video frames that can be directly illuminated by a light source. As a result, the shadow regions contain the same hue (pure color) as the background with a different intensity. Based on these properties the amount of hue-intensity difference is calculated per computer front pixel to detect and remove shadow regions in front. Then the closed front is labeled individually using a color correlogram. Finally, the HOG feature is extracted from each ROI and sent to the line SVM for personalization.

Also in the background of said topic we can state some general problems store managers are also interested in using advanced cameras to extract business intelligence information. Retailers desire available real-time Data /information about customer traffic patterns, queue lengths, and check-out waiting times to improve operational efficiency and customer satisfaction. [42] The invention is primarily involved with the use of automated and/or semi automated higher level video analysis techniques for discerning patterns of interest in video streams. The invention is directed to identifying unique patterns of interest in indoor settings. Change detection and crowding/congestion density estimation are two sub-tasks in an effective subway advanced monitoring video system. [42]

For example: people counting, crowdedness detection, any anomalous presence of persons onto the real-time track and user/people tracking. Crowding detection in subway platforms for example is of interest for closing certain passageways, dynamically scheduling additional trains, and to improve security and passenger safety in the subway environment. [42]

A video analysis method according to the invention decomposes the unique video analysis problem into two steps. A change detection algorithm is used to distinguish a background scene from a foreground. This may be done using a discontinuity preserving Markov Random Field-based approach where Data/ information from different sources (background subtraction, intensity modeling) is combined with spatial constraints to provide a smooth motion detection set of map. [42]

The obtained change detection open map is combined with geometric weights to estimate a measure of congestion of the observed area. The geometric weights are estimated by a geometry prototype module that takes into account the perspective of the camera. The weights are used to obtain an approximate translation invariant measure for crowding as people move towards or away from the camera. The segmentation scheme and framework of the invention satisfies quasi-real time computational constraints and deals with the motion detection problem within a real application scenario. The invention is preferably used within a specific context of visual surveillance and monitoring application, in particular, the task of crowd density estimation for real-time subway environment monitoring.

Some additional objectives of the research title [42]

1) The objective of the invention is a crowd detection intelligent camera to prevent covid-19 or any other need is a computer-interfaced advanced camera system that identifies and tracks groups of socially interrelated people.
2) The other objective of the invention is that a system can be used for example to track people as they wait in a checkout border, line, dotted line or at a service counter and the also implementation and the each recorded advanced camera frame is segmented into foreground regions containing several people.
3) The other objective of the invention is to a foreground regions are further segmented into individuals using temporal segmentation analysis and the Once an individual person is detected, an appearance model based on more color and endpoint, edge density in conjunction with a mean-shift tracker is used to recover the person's trajectory and also the Groups of

people are determined by analyzing inter-person distances over time.

4) The other objective of the invention is to make a computer program product and computer system for crowd detection and the computer system receives through an interface of user generated data records from a social media data storage component, a user generated data record comprises a text portion.
5) The other objective of the invention is to use a trained machine learning system as an indicator for crowd formation and the indicator is an output of the machine learning system in response to an input pair of associated location information and time information.
6) The other objective of the invention is to the same time, mobile devices are pervasive and the devices are equipped with advanced sensors, such as a camera, microphone, gyro, GPS, accelerometer and touch-screen readers and the Mobile devices provide excellent coverage of spaces of interest, and they are mobilized around the sensitive areas by people.
7) The other objective of the invention is to a large degree, the mobile devices within the premises of an enterprise satisfy a trust relation and it can be statistically safe to assume that nearby devices are to be trusted for the purpose of collaboratively mining and calibrating sensor data.

Proposed framework for blocking handling and human detection

In this segment the proposed framework has been defined in detail. The proposed framework includes six major degrees: (1) changing from RGB to gray and HSI, (2) Subtracting historical past, (three) removing shadow regions, (four) Labeling, blocking handling and filtering, (5) Extracting HOG features and (6) classification. Fig. 1 suggests the proposed framework for blocking dealing with and human detection.

A. changing from RGB to gray and HSI

The RGB body is transformed to grayscale and HSI frame. The grayscale and HSI body is used for heritage subtraction and shadow remove manners respectively.

B. Subtracting historical past

Instead of representing all of the pixel values through the same dispersion, values of each pixel are modeled as an aggregate of Gaussians to describe numerous backgrounds. based totally at the consistency and the variance of every Gaussian dispersion, the framework comes to a decision for foreground pixels. At any given time the record of a selected pixel, (x0, y0) is known as (1).

$$\{V_1,\ldots,V_t\} = \{F(x_0, y_0, m): 0<m\leq t\} \qquad (1)$$

where F is a series of frames and Vt represents the cost of a pixel density during t. Previous pixel events are represented in the form of a combination of J Gaussian dispersions. The average cost of obtaining a Vt on representation is provided in (2).

$$P(Vt) = \Sigma Ji{=}1\ wi,t * N(Vt,\mu i,t,\ \Sigma i,t) \qquad (2)$$

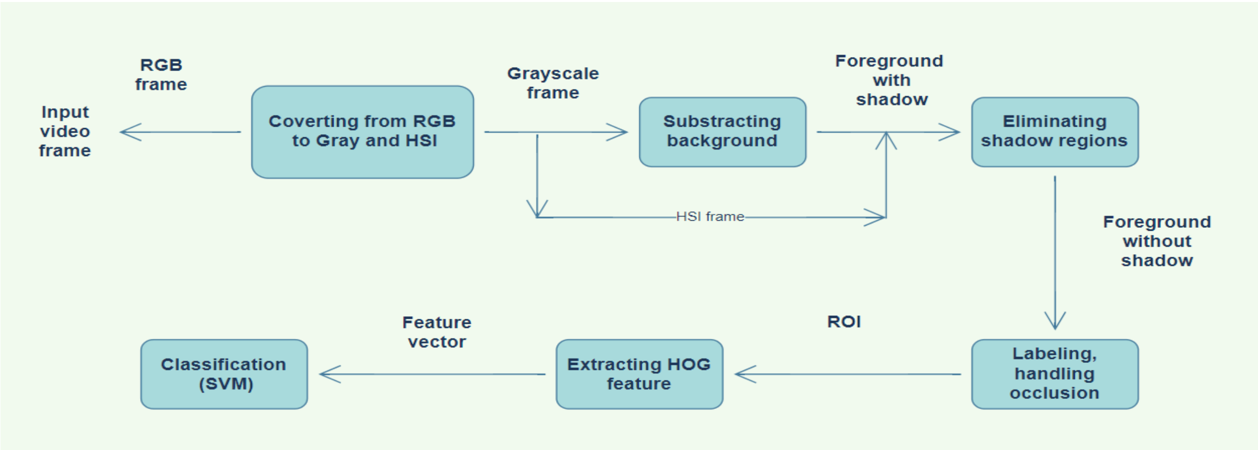


Figure 1. The proposed framework for block handling and human detection

Where J denotes the quantity of Gaussian dispersions. $\Sigma_{i,t}$, $\mu_{i,t}$ and $w_{i,t}$ denote covariance matrix, mean and weight at time t of i^{th} Gaussian respectively.
And ό represent the probability density of Gaussian dispersion. The ঌ dispersions are ranked by w_j / σ_j and first B dispersions are considered as background representation which is presented in (3).

$$\boldsymbol{B = argmi\ nb(\Sigma i{=}1b\ wi > Th)} \qquad \textbf{(3)}$$

Where ċh denotes the marginal allocation of the background representation. Foreground detection is accomplished by identifying pixels corresponding to the foreground. A pixel is considered foreground pixel if the pixel intensity value ৢV_t is beyond 2.5*σ from all ঞ dispersions. However, if ৢV_t is within 2.5 standard deviations of one or more dispersions, then the dispersion with the highest ranking i.e. w_j / σ_j value is updated using (4).

$$\boldsymbol{wJ,t,\square = (1 - a)wJ,t + a} \qquad \textbf{(4)}$$

If ৢV_t is not within 2.5* σ of any dispersion, then the dispersion with the lowest ranking replaced with a new one with ৢV_t as mean.

C. Eliminate shadow regions
The accuracy of the ROI structure depends on the output direct pre-release. Like the shadows of the object keep following the item, the background removal process considers these shadows as before. Outside, these are shadows and retain the geometric structures of the object as a result; those shadows can be misinterpreted as human. Finding shaded regions Medium value Hue-Intensity (makinga) the background and current frame of all pixels are calculated.

D. Labeling, closing and filtering
From the FWS image the frame detects closing events. A closing event is defined as, if the number of the largest binary item (BLOB) in the previous frame is greater than the number of the current BLOB in the current frame and one

BLOB in the current frame that exceeds more than one BLOB in the previous frame. After receiving the occlusion event the framework labels each BLOB in a group by computerizing the possibilities of each pixel belonging to a specific BLOB using the back-projection histogram and color correlogram. Figure 3 shows the processing model of the closure handling process.

After properly labeling the collected material the morphological closure function is applied to remove the front holes. Then the labeling of the connected part and filter is used to find the ROIs and exclude non-human circuits. Next, the framework considers the conditions associated with the human body that must be completed with a labeled item to be considered as ROI. Filter conditions are the aspect ratio and the strength of the label object.

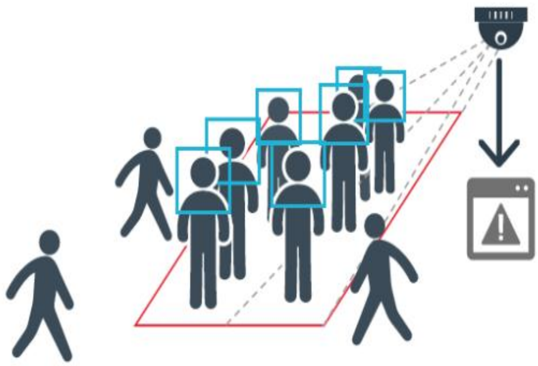


Figure 2. Processing example of block handling. [11]

E. It removes the HOG element

To exclude the Histogram of Oriented Gradients (HOG) feature [11], each ROI is adjusted to a size of 128 × 64 pixels. Then, the gradient and magnitude are subtracted from each ROI by making a convolution with a horizontal and vertical kernel represented by [-1 0 1] and [-1 0 1] T respectively. Next, the gradient image is separated by 8 × 8 pixel cells. In each cell the histogram is calculated by taking a sample of gradient shape (0 ° -180 °) into 9 barrels of equal size. Each barrel represents the size of the corresponding shape. After producing a histogram for each cell, 2 × 2 cells are grouped into blocks with 50% spacing to make the illumination element constant. Then, all block histograms are combined to produce a vector element. Finally, the element vector is common in L2-normal to produce a HOG element.

F. Separation

Finally, the HOG vector feature is sent to the SVM line for human detection. SVM is a category of supervised genes. In the training database collected by clusters, SVM line aims to detect marginal-margin hyperplane, resulting in very large divisions between groups.

Effects of test

In this section, the results of the proposed closure and retrieval management framework are described. The test was performed on an Intel Core i5 3.20 GHz CPU and 4 GB of RAM memory using MATLAB space. Videos captured with a still camera with an average of 25 fps and a resolution of 320 × 240 pixels (QVGA) in urban, urban and rural areas. Most people in the data set are standing or walking. Other cases of partial closure occur, including people walking through certain objects or another person. The proposed framework is trained in 140 frames and tested in 1280 frames

TABLE I. PRECISION AND RECALL VALUE AT DIFFERENT ENVIRONMENT CONDITIONS

Frame type	Environmental conditions	Total frame	Prec (%)	Rec (%)	Avg. time(s)
Indoor	Normal condition, uneven illumination and low contrast	433	94.7	93.4	0.46
Outdoor		615	93.2	92.8	
Complex Background	Normal illumination	232	91.3	89.3	
	Average	1280	93.1	91.8	

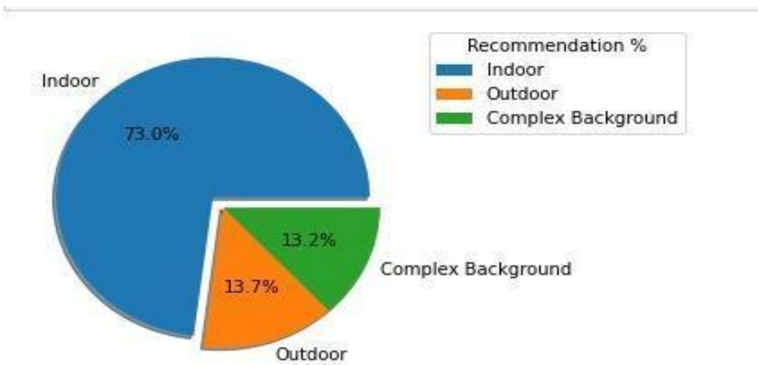
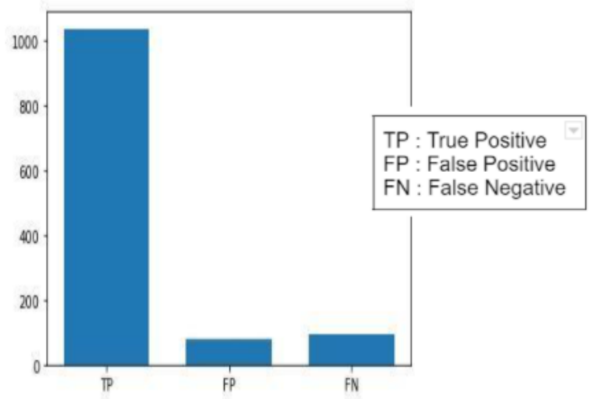


TABLE II. COMPARISON OF RESULTS BETWEEN PROPOSED FRAMEWORK AND [2]

Framework	TP	FP	FN	Prec (%)	Rec (%)
The proposed framework	1039	78	92	93.1	91.8
[2]	754	116	312	86.7	70.3



Prec and Recognition values are calculated electronically from different types of video frames that are filmed in different environments and lighting conditions. Table I shows the accuracy and value of memory in a variety of environmental conditions. In the table TP, FP and FN mean True Positive, False Positive and False Negative respectively. The proposed framework demonstrates high responsiveness to internal and external video frames and provides satisfactory results for video frames that contain complex backgrounds. The results from the proposed framework are compared with [4].

Conclusion

This paper proposes a framework for managing the closure and acquisition of a person, with the aim of getting people out of a continuous framework and high flexibility. Initially, the RGB framework is converted to a gray framework and an HSI framework. Then the back removal is done to remove the front circuits. After that, the shadow removal process is used to remove the shadow regions from the front in order to obtain an accurate ROI. Then the label is applied using a color correlogram to treat occlusion and filtering is used to remove the sounds. Finally, the HOG feature vector is extracted from the ROI and sent to the line SVM for human location. The proposed framework has limited access to videos provided by the still camera. This framework may not produce the best results if a small portion of the closed person is exposed. This work will also be extended to recruited people from the best areas. It will also focus on using a segment-based acquisition to better manage closure.

References

[1] Y. Ma, L. Deng , X. Chen and N. Guo, “Integrating Orientation Cue With EOH-OLBP-Based Multilevel Features for Human Detection, ” IEEE Trans. Circuits Syst. Video Technol., vol. 23, no. 10, pp. 1755 - 1766, Oct. 2013.

[4] B. Jun, I. Choi, D. Kim, 'Local Transform Features and Hybridization for Accurate Face and Human Detection," IEEE Trans. Pattern Recognit. Mach. Intell., vol. 35, no. 6, pp. 1423 - 1436, June 2013.

[3] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, “Object detection with discriminatively trained part based models,” IEEE Trans.

Pattern Recognit. Mach. Intell., vol. 32, no. 9, pp. 1627– 1645, Sep. 2010.

[4] R. Tong , D. Xie and M. Tang, “Upper Body Human Detection and Segmentation in Low Contrast Video, ” IEEE Trans. Circuits Syst. Video Technol., vol. 23, no. 9, pp. 1502 - 1509, Sept. 2013.

[5] B. Wu and R. Nevatia, “Detection and segmentation of multiple, partially occluded objects by grouping, merging, assigning part detection responses,” Int. J. Comput. Vis., vol. 82, no. 2, pp. 185–204, Apr. 2009.

[6] M. Jalali Moghaddam, E. Shaabani, and R. Safabakhsh. Crowd density estimation for outdoor environments. In Proceedings of the 8th International Conference on Bioinspired Information and Communications Technologies, BICT ’14, pages 306–310, ICST, Brussels, Belgium, Bel- gium, 2014. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[7] D. Kim, Y. Lee, B. Ku, and H. Ko. Crowd density estimation using multi-class adaboost. In Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on, pages 447–451, Sep 2012.

[8] R. Ma, L. Li, W. Huang, and Q. Tian. On pixel count based crowd density estimation for visual surveillance. In Cybernetics and Intelligent Systems, 2004 IEEE Conference on, volume 1, pages 170–173 vol.1, Dec 2004.

[9] W. Ma, L. Huang, and C. Liu. Advanced local binary pattern descriptors for crowd estimation. In Computational Intelligence and Industrial Application, 2008. PACIIA ’08. Pacific-Asia Workshop on, volume 2, pages 958–962, Dec 2008.

[10] W. Ma, L. Huang, and C. Liu. Crowd estimation using multi-scale local texture analysis and confidence-based soft classification. In Intelligent Information Technology Application, 2008. IITA ’08. Second International Symposium on, volume 1, pages 142–146, Dec 2008.

[11] Mandar Ganesh Sohani, Dr. S.A. Patekar, “A Survey on Human detection in Crowd Density Estimation for Video Surveillance”, International Journal of Mechanical Engineering., ISSN: 0974-5823 Vol. 6 No. 3 December, 2021

[12] H.Rahmalan, M.S.Nixon, and J.N.Carter, “On crowd density estimation for surveillance”, in Proc. Inst. Eng. Technol. Conf. Crime security, pp.540-545, 2006.

[13] X.Li, L.Shen, and H.Li, “Estimation of crowd density based on wavelet and support vector Machine”, Trans. Inst. Meas.Control, vol.28, no.3, pp. 299-308, Aug. 2006.

[14] D.Kong, D.Gray, and T.Hat, “A viewpoint invariant approach for crowd counting”, in Proc.Int. Conf. Pattern Recog, pp. 1187-1190, 2006.

[15] Volker Eiselein, Hajer Fradi, Ivo Keller, Thomas Sikora, Jean-Luc Dugelay, “Enhancing Human Detection using Crowd Density Measures and an adaptive Correction Filter”, 2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance

[16] S. D. Khan, S. Bandini, S. Basalamah, and G. Vizzari, “Analyzing crowd behavior in naturalistic conditions: Identifying sources and sinks and characterizing main flows,” Neurocomputing, vol. 177, pp. 543–563, Feb. 2016.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Proc. Adv. Neural Inf. Process. Syst., 2012, pp. 1097–1105. [18] W. Kuo, B. Hariharan, and J. Malik,

"DeepBox: Learning objectness with convolutional networks," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2015, pp. 2479–2487.

[19] V. Lempitsky and A. Zisserman, "Learning to count objects in images," in Proc. Adv. Neural Inf. Process. Syst., 2010, pp. 1324–1332.

[20] W. Li, H. Li, Q. Wu, F. Meng, L. Xu, and K. N. Ngan, "Headnet: An endtoend adaptive relational network for head detection," IEEE Trans. Circuits Syst. Video Technol., to be published.

[21] J. Liu, C. Gao, D. Meng, and A. G. Hauptmann, "Decidenet: Counting varying density crowds through attention guided detection and density estimation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2018, pp. 5197–5206.

[22] L. Liu, H. Wang, G. Li, W. Ouyang, and L. Lin, "Crowd counting using deep recurrent spatial-aware network," Jul. 2018, arXiv:1807.00601. [Online]. Available: https://arxiv.org/abs/1807.00601

[23] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in Proc. Eur. Conf. Comput. Vis. (ECCV). Amsterdam, The Netherlands: Springer, 2016, pp. 21–37.

[24] D. Onoro-Rubio and R. J. López-Sastre, "Towards perspective-free object counting with deep learning," in Proc. Eur. Conf. Comput. Vis. Springer, 2016, pp. 615–629.

[25] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2014, pp. 1717–1724.

[26] V.-Q. Pham, T. Kozakaya, O. Yamaguchi, and R. Okada, "Count forest: CO-voting uncertain number of targets using random forest for crowd density estimation," in Proc. IEEE Int. Conf. Comput. Vis., Dec. 2015, pp. 3253–3261.

[27] V. Rabaud and S. Belongie, "Counting crowded moving objects," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR), vol. 1, Jun. 2006, pp. 705–711.

[28] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in Proc. Adv. Neural Inf. Process. Syst., 2015, pp. 91–99.

[29] D. B. Sam, S. Surya, and R. V. Babu, "Switching convolutional neural network for crowd counting," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), vol. 1, Jul. 2017, pp. 4031–4039.

[30] M. Saqib, S. D. Khan, N. Sharma, and M. Blumenstein, "Person head detection in multiple scales using deep convolutional neural networks," in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2018, pp. 1–7.

[31] M. Saqib, S. D. Khan, N. Sharma, and M. Blumenstein, "Crowd counting in low-resolution crowded scenes using region-based deep convolutional neural networks," IEEE Access, vol. 7, pp. 35317–35329, 2019.

[32] M. Shami, S. Maqbool, H. Sajid, Y. Ayaz, and S.-C. S. Cheung, "People counting in dense crowd images using sparse head detections," IEEE Trans. Circuits Syst. Video Technol., to be published.

[33] Z. Shen, Y. Xu, B. Ni, M. Wang, J. Hu, and X. Yang, "Crowd counting via adversarial cross-scale consistency pursuit," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 5245–5254.

[34] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” Sep. 2014, arXiv:1409.1556. [Online]. Available: https://arxiv.org/abs/1409.1556 [35] V. A. Sindagi and V. M. Patel, “Generating high-quality crowd density maps using contextual pyramid CNNs,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 1861–1870.

[36] H. Ullah, A. B. Altamimi, M. Uzair, and M. Ullah, “Anomalous entities detection and localization in pedestrian flows,” Neurocomputing, vol. 290, pp. 74–86, May 2018.

[37] H. Ullah, M. Uzair, M. Ullah, A. Khan, A. Ahmad, and W. Khan, “Density independent hydrodynamics model for crowd coherency detection,” Neurocomputing, vol. 242, pp. 28–39, Jun. 2017.

[38] M. Ullah, F. A. Cheikh, and A. S. Imran, “Hog based real-time multi-target tracking in Bayesian framework,” in Proc. 13th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS), Aug. 2016, pp. 416–422.

[39] Julio Cezar,Soraia Raupp Musse,Cláudio Rosito Jung “Crowd Analysis Using Computer Vision Techniques” IEEE SIGNAL PROCESSING MAGAZINE SEPTEMBER 2010

[40] Muhammad Saqib, Sultan Daud Khan, Michael Blumenstein “Texture-Based Feature Mining for Crowd Density Estimation: A Study” 978-1-5090-2748-4/16/$31.00 2016 IEEE

[41] Shayhan Ameen Chowdhury , Mohammed Nasir Uddin , Mir Md. Saki Kowsar , Kaushik Deb,”Occlusion Handling and Human Detection Based on Histogram of Oriented Gradients for Automatic Video Surveillance”978-1-5090-6122-8/16/$31.00 ,2016 IEEE

[42] Umakant Butkar and et al.“Crowd Detection Intelligent Camera to Prevent Covid-19 or any Other Need” Patent Granted (Australian Government) Patent number: 202110266816th day of March 2022.