ISSN: 0974-5823 Vol. 6 No. 3 December, 2021

International Journal of Mechanical Engineering

A Survey on Human detection in Crowd Density Estimation for Video Surveillance

Mandar Ganesh Sohani
PhD Scholar

Dr. S.A. Patekar
Guide & Professor

Department of Computer Engineering, Vidyalankar Institute of Technology, University of Mumbai

ABSTRACT

Crowds can be seen in numerous day-to-day life situations and it'll be engaging to recognize, dissect and break the challenges involved in crowd density estimation. The density of a crowd is a vital parameter in several operations like operation of crowd for safety and surveillance for law enforcement, development of public transport structure which have been divided using automated or semi-automated computer vision ways. Mortal discovery in a videotape surveillance system has vast operation areas including suspicious event discovery and mortal exertion recognition. In the current terrain of our society suspicious event discovery is a burning issue. For that reason, this paper proposes a frame for detecting humans in different appearances and acts by generating a mortal point vector. Originally, every pixel of a frame is represented as an objectification of several Gaussians and use a probabilistic system to refurbish the representation. These Gaussian representations are also estimated to classify the background pixels from focus pixels. Shadow regions are excluded from focus by exercising a Hue-Intensity difference value between background and current frame. Also morphological operation is used to remove discontinuities in the focus uprooted from the shadow elimination process. Partial occlusion running is employed by color correlogram to marker objects within a group.

Kay Indicators : RGB, HSI, Gray Image, HOG,Pixel Frames, PBA

OBJECTIVE OF RESEARCH

The objectives of my research are

1. Analyzing the existing methods used for crowd density estimation by considering their accuracy for achieving second and third objective that are bellow
2. Designing a method to estimate crowd density by counting the number of people in a scene.
3. Increasing accuracy by blending various methods so that estimation will be more & more closer to actual

INTRODUCTION OF RESEARCH IDEA:

Crowd counting from unconstrained scene images is a pivotal task in numerous real- world operations like civic surveillance and operation, but it's greatly challenged by the camera's perspective that causes huge appearance variations in people's scales and reels[22].

As one of the most important contents in intelligent videotape surveillance, the crowd viscosity estimation plays a central part in public safety, crowd operation, business control and so on. Since Davies etal.[3] presented an automatic estimation system for crowd viscosity using image processing in 1995, numerous styles have been proposed. Still, there are still several complex challenges which make this subject a focus of exploration. Expansive reviews on crowd viscosity estimation and applicable subjects can be planted. We roughly classify the being styles into two orders, of which one is the styles grounded on holistic point birth ways, and the other is the styles grounded on individual discovery ways [10].

The escalation of computer vision usages impelled mortal discovery as an active exploration field. Mortal discovery in a videotape surveillance system has vast operation areas including mortal locomotion characterization, fall discovery for cases and intelligent gestural stoner interface (wiimote, kinect, smart Television). Mortal discovery is a deep- seated and demanding issue because of two challenges

1) HumansIntra-class divergences like appearance, apparel, skin color and disguise;

2) External issues like uneven illumination and cluttered background.

Current mortal discovery fabrics can be perished into two processes. One process employs a sliding window, while the other process employs a part- grounded discovery. The sliding window grounded process can be bettered in two areas: Composing more sapient features to ameliorate discovery rate and use effective training styles to learn better classifiers. Extensively used features involve Haar sea, Overeater, shapelet, edge exposure histogram (EOH), edgelet, region covariance and LBP [23].This paper proposes a frame to describe clotted humans by rooting focus from background using background deduction process. The main emphasis of this paper is to exclude shadow regions from focus to find accurate ROI. Murk can be defined as a portion of regions in a videotape frame that aren't directly illuminated by light source. As a result, shadow regions contain same tinge (pure color) as the background with different intensity values. Grounded on these parcels a tinge- intensity difference value is reckoned for every focus pixel to descry and exclude shadow regions in focus. Also clotted centers are labeled collectively by exercising color correlogram. Eventually, Overeater point is uprooted for each ROI and transferred to direct SVM for mortal discovery. [40]

LITERATURE REPORT

Paper 1: Julio Cezar Et.al states in his paper Crowd Analysis Using Computer Vision Techniques

The behavioral analysis of human crowds using computer vision algorithms is, and probably will be for a long time, the focus of attention for several researchers due to the possible potential applications. This problem presents challenges of great complexity that could involve researchers from several areas and backgrounds. In particular, the integration of computer vision and computer graphics is becoming more popular in both crowd analysis and synthesis[39].

Paper 2: Muhammad Saqib Et.al States in their research paper Texture-Based Feature Mining for Crowd Density Estimation: A Study has evaluated different texture features for crowd density estimation and count. We also proposed a two stage classification and regression framework. In classification, small blocks of crowd are classified as very low, low, medium and high density. At the second stage, Gaussian Process Regression is used to regress features to count[40].

Paper 3: Shayhan Ameen Chowdhury Et.al evaluated in their researcher Occlusion Handling and Human Detection Based on Histogram of Oriented Gradients for Automatic Video Surveillance, proposed a framework for occlusion handling and human detection, with the goal to detect humans from continuous frame sequences with higher adaptability. Initially, the RGB frame is converted to grayscale and HSI frame. Then background subtraction is performed to extract foreground regions. After that, the shadow elimination process is used to remove shadow regions from foreground to find the accurate ROI. Then labeling is utilized by using color correlogram for occlusion handling and filtering is employed to remove noises. Finally, HOG feature vector is extracted from ROI and sent to linear SVM for detecting human region. The proposed framework is limited to detect humans from videos provided by a stationary camera. This framework may not provide better results if small portion of an occluded human is visible. This work will be extended to detected humans from moving background. And also focus will be given to implement human part-based detection for better occlusion handling[41].

Paper 4: Tao Zhao Et.alstates in Tracking Multiple Humans in Complex Situations, described our methods for segmentation and tracking of multiple humans in complex situations and estimation of human locomotion modes and phases (coarse 3D body postures). We use explicit 3D shape model in segmentation and tracking of multiple humans. The shape model enables segmenting multiple-human with persistent occlusion (e.g., walking together) and provides a representation for tracking. Three-dimensional model combined with camera model and the assumption that people move on a ground plane makes the approach suitable for a wide range of viewpoints, automatically scales the model as people moves, facilitates occlusion reasoning, and provides 3D trajectories [1].

Paper 5: Ya-Li Hou Et.al states in his People Counting and Human Detection in a Challenging Situation, foreground pixels from both moving people and near stationary people have been considered to estimate their number. After a closing operation over foreground pixels, one can observe a linear relationship between the number of people and foreground pixels. The best estimation results, with a 10% average error, were achieved when both foreground pixels and closed foreground pixels are learned in a neural network[4].

Paper 6: Bingyin Zhou Et.al states in his research named as Computer Vision and Image Understanding, introduced a new method to estimate the crowd density based on the combination of HOSVD and SVM. We treat images as higher-order tensors, and the density feature vectors are extracted using their projections onto the principal tensor subspace. A multi-class SVM classifier is designed to classify the feature vectors into different density levels. Experimental results show that the accuracy of our method achieves 96.33%, in which the misclassified images are all concentrated in their neighboring categories[10].

Paper 7: Volker Eiselein Et.al proves some methods in Enhancing Human Detection using Crowd Density Measures and an adaptive Correction Filter, present a strategy of exploiting crowd density information to enhance human detection. By means of automatically estimated crowd density maps, the detection threshold of a human detector can be adjusted according to the scene crowd context. In order to cope with false positive detections of wrong size, a dynamically-learning correction filter exploiting the aspect ratio of detections is proposed[15].

Paper 8: Lingbo Liu Et.al states in his research title Crowd Counting using Deep Recurrent Spatial-Aware Network, e introduce a novel Deep Recurrent SpatialAware Network for crowd counting which simultaneously models the variations of crowd density as well as the pose changes in a unified learnable module. It can be regarded as a general framework for crowd map refinement. Extensive experiments on four challenging benchmarks show that our proposed method achieves superior performance in comparison to the existing state-of-the-art methods. In our future research, we plan to delve into the research of incorporating our model in other existing crowd flow prediction frameworks[22].

PROPOSED FRAMEWORK FOR OCCLUSION HANDLING AND HUMAN DETECTION

In this section the proposed frame has been described in details. The proposed frame consists of six main stages (1) Converting from RGB to Gray and HSI, (2) Subtracting background, (3) Eliminating shadow regions, (4) Labeling, occlusion handling and filtering, (5) Extracting HOG features and (6) Classification [41].

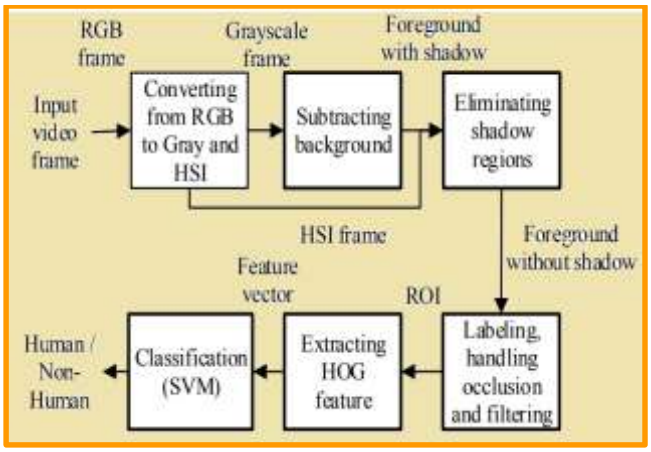

Fig. 1 Proposed frame for occlusion running and mortal discovery.

A. Converting from RGB to Gray and HSI

The RGB frame is converted to grayscale and HSI frame. The grayscale and HSI frame is used for background deduction and shadow exclude process independently.

B. Subtracting background

Rather of representing all the pixel values by same dissipation, values of each pixel are modeled as an admixture of Gaussians to describe multitudinous backgrounds. Grounded on the thickness and the friction of each Gaussian dissipation, the frame decides focus pixels. At any given time the history of a specific pixel.

C. Eliminating shadow regions

The delicacy of ROI construction relies on generating accurate focus birth. As murk of an object

continually follow the object, background deduction process considers these murk as focus. Beside, these murk also save the geometric parcels of an object as a result; those murk can be misclassified as mortal. For detecting shadow regions a Hue-Intensity difference value between background and current frame for every pixel is calculated.

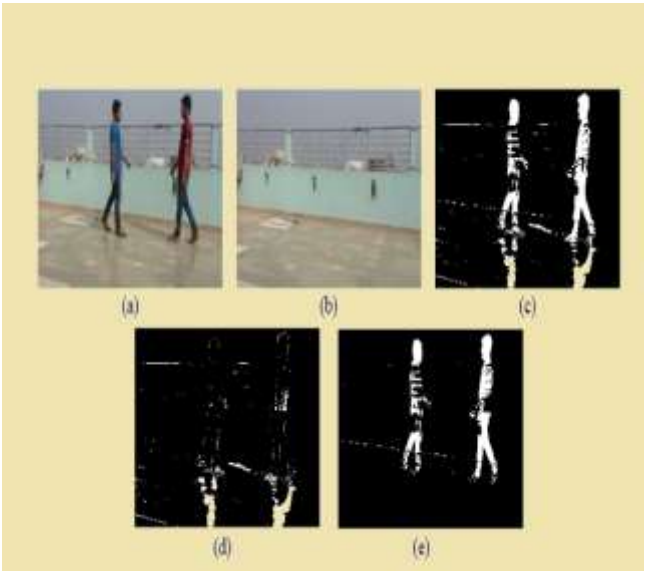

Fig. 2. shadow elimination process:

(a) current frame, (b) background frame, (c) foreground with shadow image, (d) shadow region and (e) foreground without shadow image.

(b)

D.Labeling, handling occlusion and filtering

From the FWS image the frame detects occlusion events. An occlusion event is defined as, if double large object (BLOB) number in the former frame is lesser than the BLOB number in the current frame and one of the BLOBs in current frame overlaps with further than one BLOBs in the former frame. After detecting an occlusion event the frame marker individual BLOB in a group by calculating liability of each pixel belonging to a particular BLOB with the application of back- protuberance histogram and color correlogram. Fig. 3 shows the processing illustration of occlusion running process After rightly labeling grouped objects morphologically.

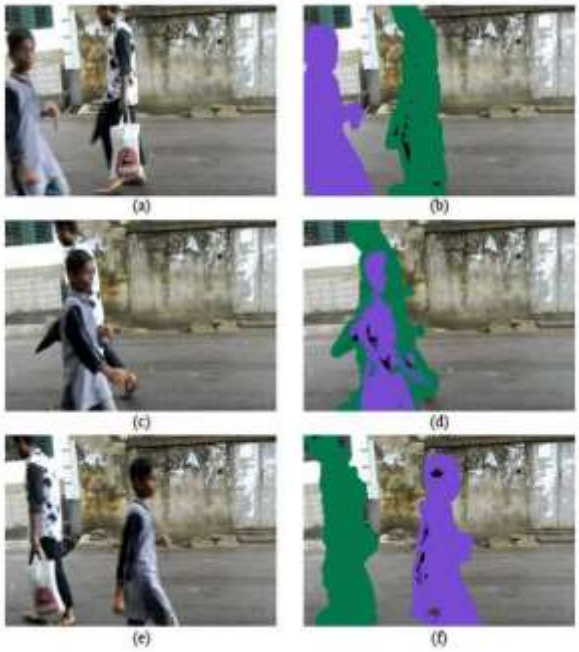

Fig. 3 Processing illustration

Fig. 3. Processing occlusion handling. Left and right column represent input frame and labeled frame respectively. (a) & (b) before occlusion, (c) & (d) during occlusion and (e) & (f) after occlusion.

E. Extracting HOG feature

To extract Histogram of Oriented Gradients (HOG) [11] feature, each ROI is resized into 128×64 pixels. Then, gradient orientation and magnitude is extracted for each ROI by performing a convolution with a horizontal and vertical kernel represented by [-1 0 1] and [-1 0 1]T respectively. After that, the gradient image is divided into cells of 8×8 pixels. For each cell a histogram is computed by sampling the gradient orientation (0°-180°) into 9 equal size bins. Each bin represents the magnitude of the corresponding orientations. After generating histogram for each cell, 2×2 cells are grouped into blocks with 50% overlap to make the feature illumination invariant. Then, all block histograms are concatenated to generate feature vector. Finally, the feature vector is normalized with L2-norm to generate HOG feature. The size of HOG feature is 9×7×15×4=3780 [41] from base paper.

F. Classification

Eventually, the Overeater point vector is transferred to a direct SVM for mortal discovery. SVM is a supervised periphery classifier. For two grouped training dataset, direct SVM intends to find a maximum- periphery hyperplane, which leads to largest separation between the groups.

SOME IMPORTANTS ANALYSIS DURING STUDY

The crowd analysis can be performed for three different purpose

1. People Counting
2. People Tracking

3. People Behavior Understanding

The people counting is prime area of interest for this research work. The people counting methods can be mainly classified into three different types

1. Pixel Based Analysis
2. Texture Level Analysis

PIXEL-BASED ANALYSIS

PBA styles depend on features similar to individual pixel analysis attained through background deduction models or edge discovery to estimate the number of people in a scene. Since veritably low- position features are used, this class is substantially concentrated on viscosity estimation rather than precise people counting. There's a direct relationship between number of focus pixels and number of people in a scene. The focus image area, border, and edge pixels like pixel features are useful to prognosticate the viscosity. Still, the focus image segmentation algorithm isn't ideal and needs to correct the weight of uprooted pixels due to the impact of perspective deformation. Thus pixel statistical algorithms have bad results in high viscosity crowds. It can be used to give the original vaticination of crowd viscosity, and estimate the crowd viscosity as low viscosity or high viscosity.

TEXTURE LEVEL ANALYSIS:

This class explores advanced- position features when compared to pixel- grounded approaches, it's also substantially used to estimate the number of people in a scene rather than relating individualities. Different viscosity crowds have different texture patterns for texture analysis. Images of low viscosity crowds show coarse texture, while images of high viscosity crowds show fine texture.

Texture analysis can be performed by using four styles of argentine- position dependence matrix, straight lines parts, Fourier analysis, and fractal dimension. After analysis, crowd viscosity can be estimated using neural network, statistics (Bayesian), and a befitting function- grounded approach.

FOCUS OF STUDY:

After landing an image frame from videotape, the Gaussian Mixture Model (GMM) can be applied to prize focus information. Gaussian admixture model for adaptive background modeling, which models each pixel as a admixture of Gaussians to determine whether or not a pixel is part of the background. The system can separate the background and focus effectively and can be used for real-time shadowing. In proposition, any complex scene can be generated by a certain number of Gaussian distributions. Although the GMM can overcome interferences of illumination and murk to some extent, the situation of mis gauging the focus as the background is necessary because the target movement speed is too slow or the color of focus is analogous to the background.

The uprooted focus image can correspond of interferences similar as cracks, depressions and star points, which have bad goods on focus edge birth. It'll be enhanced by barring noise (star points) using Median filtering. The morphological operation dilation is used to remove the crack followed by erosion operation to keep the original boundary shape features of the focus.

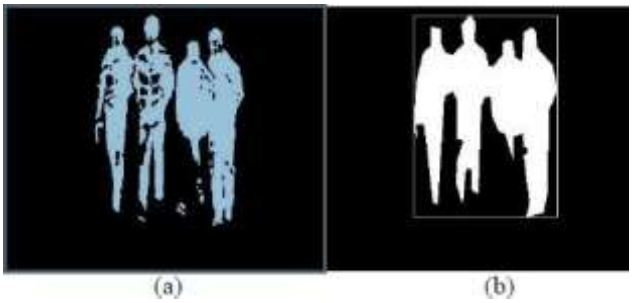

Fig 4: Foreground extracted a)Before enhancement b) After Enhancement

1. SETTING REGION OF INTEREST:

The extracted foreground may consists of abnormal projections, especially in the large scale monitoring. The effect of abnormal projections is due to perspective of camera location. Theforeground region area size will be different at different distance from camera. Therefore the image is divided into sub regions. Each sub region is considered as separate area of interest.

Fig 5: Image divided in to sub-regions

2. INITIAL PREDICTION OF CROWD DENSITY:

The crowd density can be initially predicted by using region area as follows

$N_k / A \leq T_k$

Where, N_k is the total number of foreground pixels in k^{th} sub-region A is the area of k^{th} sub-region (width * height)

Tk is threshold set for k^{th} sub-region

If the k^{th} sub-region satisfies the equation it means the crowd density is low and we can count the number of people by using the Pixel statistical model. But if the equation is not satisfied it means the crowd density is high and we will use a texture analysis method.

3. PIXEL STATISTICAL MODEL:

For a particular region, the number of people in this region can be calculated according to the corresponding fitting straight line. The fitting straight line is trained by a method of linear regression. According to the number of foreground pixels and count the true number of people in this region artificially for each region of training samples, using the least squares method, fitting out four straight lines that the number of pixels corresponds to the number of people.

REFERENCES:

[1] T.Zhao and R.Nevatia,"Tracking multiple humans in complex situations", IEEE Trans, Pattern Anal.Mach. Intell., vol.26, no.9, pp.1208-1221, sep.2004.

[2] T.Zhao, R.Nevatia, and B.Wu, "Segmentation and tracking of multiple humans in crowded Environments", IEEE Trans.Pattern Anal.Mach.Intell.,vol30,no.7,pp.1198-1211,Jul.2008

[3] Y. J., V. S., and A. Davies, "Image processing techniques for crowd density estimation using a reference image," in ACCV, vol. 3, 1995, pp. 6–10.

[4] Ya-Li Hou, and Grantham K.H.Pang, "People counting and human detection in a

challenging situation", IEEE Trans.sys.man and cybernetics.,vol.41,No.1, pp.24-33, Jan.2011.

[5] V.Rabaud and S.Belongie, "Counting crowded moving objects", in Proc.IEEE Conf.Comput Vis.Pattern Recog.,pp.705-711,2006

[6] S-Y.Cho,T.W.S.Chow, and C-T.Leung, "A neural-based crowd estimation by hybrid global learning algorithm", IEEE Trans.Syst.Mcn Cybern.B.Cybern., vol.29.no.4,pp.535-541, Aug.1999.

[7] R.Ma,L.Li,W.Huang, and Q.Tian,"On pixel count based crowd density estimation for visual Surveillance", in Proc.IEEE conf.Cybern.Intell.Syst.pp.170-173,2004.

[8] P.Karpagavalli, A.V.Ramprasad, "Estimating the density of the people and counting the Number of people in a crowd environment for human safety", IEEE International conference on Communication and Signal Processing, 2013

[9] P.Kilambi, O.Masoud, and N.Papanikolopoulos, "Crowd analysis at mass transit site", in Proc.IEEE Intell. Transp. Syst.Conf.,pp.753-758, 2006.

[10] A.N.Marana, S.A.Velastin, L.F.costa, and R.A.Lotufo, "Estimation of crowd density using image processing", in Proc.IEEE colloq.Image Process.Security Appl.,pp.11/1-11/8,1997

[11] A.N.Marana, L.Da Fontoura costa, R.A.Lotufo, and S.A.Velastin, "Estimating crowd Density with Minkowski fractal dimension", in Proc. Int. Conf. Acoust., speech, Signal Process.,pp.3521-3524,1999.

[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 end-toend 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