

# Robust Approach of Edge detection in Videos Using Spatial-Temporal Features

Abdullah Jamaan Alzahrani<sup>1</sup> and Jasim Khan<sup>2</sup>

<sup>1</sup> Department of Computer Science and Software Engineering, University of Hail  
Hail, Saudi Arabia

<sup>2</sup> Siemens Italy

## Abstract

Edge detection is still a challenging problem and researchers are focusing to investigate this problem using different techniques. Edge detection is an important preprocessing step in most of image processing applications. The application ranges from real-time video surveillance, traffic surveillance to medical imaging applications. Current state-of-the-art methods for edge detection are filter based and do not incorporate spatial-temporal information among the consecutive frames. We propose a robust approach for edge detection by exploiting spatial temporal information that possess an important cue to robust edge detection. This is achieved by extracting hybrid features in terms of pairwise local binary pattern (P-LBP) and scale invariant feature transform (SIFT). These features are used to train an MLP neural network during the training stage, and the edges are inferred from the test videos during the testing stage. The experimental evaluation is conducted on a benchmark dataset commonly used for edge detection.

**Keywords:** *Neural Networks, Local Binary Pattern, Edge Detection, and Scale Invariant Feature Transform.*

## 1. Introduction

Edge detection is a significant component in many applications including but not limited to video surveillance [15], [16], [17], [18], ambient intelligence [19], [20], [21], [22], human-computer interaction systems [27], [28], [29], [30], [31], and health-care [23], [24], [25], [26]. In spite of incredible research efforts and many encouraging advances in the last ten years, accurate detection of edges is still a challenging problem. Information about the edges in the image can be used in image segmentation [45], [47], [48], [51] and identification of different shapes [32], [33], [34], [46]. Typically, edge detection algorithms are based on the notion that any discontinuity in the gray level value will lead to the edges. Edge can be referred to those points in the image that has sharp boundary between the object and background. The most common and important application of edge detection is image segmentation. In image segmentation problem, image is divided into meaningful homogeneous blocks [37], [38], [40]. Efficient image segmentation involve the following two important steps: 1)

Clustering all the pixels that satisfy the criteria of homogeneity. 2) Once the homogeneous regions are detected the second step is to find the boundaries between different homogeneous regions. This is where the role of edge detection comes in.

To improve edge detection, many techniques have been proposed and they rely on differential geometry [1], anisotropic Gaussian kernel [2], single pixel imaging [3], and adaptive approach [4]. Other approaches include difference filter [5], canny edge detector [6], laplacian of Gaussian [7] method. In canny edge detector, we first apply Gaussian filter to smooth the image in order to remove noise. Then in the next step, the gradient of image ( $G_x$  and  $G_y$ ) is computed. After computing the gradients non-maxima suppression is employed to suppress low gradient pixels and keep high gradient edges. Some other methods employ entropy for the edge detection for example [8], [9], [10].

Artificial Neural network (ANN) proves to provide a robust solutions in many classification problems [35], [36], [39]. Keeping in the view, the benefits of ANN, a method is proposed in [11] that trained a network that can detect edges in the images. The ANN is more useful because multiple inputs and outputs can be used during the training. However, designing effective features for edge is difficult due to large intra-class variation arising from pose appearance [49] and temporal variations [50]. Therefore, it is crucial to design discriminative features [42], [44]. It is worth noticing that the combination of features could boost the discriminative power of a model. Considering combination of two features could provide much more information than observing occurrence of two features individually.

We propose a robust approach to combine different spatio-temporal features [41], [43] to design a unified model for edge detection. For this purpose, we consider pairwise local binary pattern (P-LBP) [12] and scale invariant feature transform (SIFT) [13]. The P-LBP and SIFT are spatial and temporal features, respectively. We then adopt an MLP neural network using the spatio-temporal features

(P-LBP and SIFT) during the training stage [52], [53], [54]. The edges are inferred from the testing videos during the testing stage. The overall process of our proposed approach is presented in Fig. 1. The rest of the paper is organized as follows: Section 2 presents the related work; Section 3 elaborates our proposed method of features extraction; Section 4 presents experimental evaluation; and Section 5 concludes this paper.

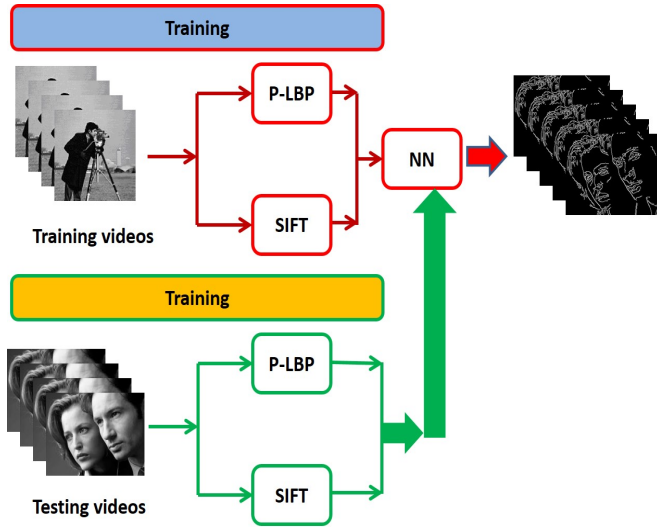


Figure 1. Flow diagram. The training videos are provided to train the neural network and testing videos are used to recognize edges during the testing stage.

## 2. Proposed Methodology

First we consider spatial feature P-LBP. In order to calculate pairwise rotation invariant local binary pattern (P-LBP), for each pixel on the video frame, a threshold is applied to its symmetric neighbor set on a circle of radius  $R$  and the result is considered as a binary number. The P-LBP is formulated in Eq. (1) and Eq. (2).

$$\Psi(x) = \sum_{p=0}^{P-1} s_p(x) 2^p \quad (1)$$

$$s_p(x) = \begin{cases} 1, & V(N_p(x)) \geq V(x) \\ 0, & V(N_p(x)) < V(x) \end{cases} \quad (2)$$

Where  $x$  is a coordinate,  $N_p(x)$  is the  $p$ th neighbor of point  $x$ , and  $V(x)$  is the pixel value of point  $x$ . It is worth noticing that the function  $s_p(x)$  isn't affected by changes in mean luminance. Therefore, the P-LBP could achieve invariance. In order to achieve rotation invariance, Ojala et al. [39] formulates the rotation invariance in Eq. (3)

$$\Psi^r(x) = \min\{ROR(\Psi(x), i) \mid i \in [0, P-1]\} \quad (3)$$

Where  $ROR(x, i)$  calculates a circular bit-wise right shift for  $i$  times on  $P$ -bit number  $x$ . Ojala et al. [39] also investigate that patterns presenting limited spatial transitions indicate the fundamental properties of frame microstructure. For example, the pattern represented by the sequence 11110000 describes a local edge, and the pattern represented by the sequence 11111111 describes a flat region or a dark spot. To formally define these patterns, a uniformity measure is formulated in Eq. (4),

$$U(x) = \sum_{p=1}^P |s_p(x) - s_{p-1}(x)| \quad (4)$$

where  $s_P(x)$  is defined as  $s_0(x)$ . The uniform patterns is subject to the condition  $U(x) \leq 2$ .

We calculate temporal feature using SIFT. Scale invariant feature transform (SIFT) have different properties that make them suitable for extracting temporal information. In fact, SIFT features are invariant to scaling and rotation, and partially invariant to change in illumination and 3D camera viewpoint [55], [56]. They are well concentrated in both the spatial and frequency domains, reducing the chances of disruption by occlusion and other unprecedented noise. SIFT features are very distinctive, which allows a single feature to be uniquely identified, providing a basis for human action recognition [57], [58], [59].

For extracting the SIFT features, the scale space of a video frame is defined as a function  $L(x, y, \sigma)$  that is produced from the convolution of the input video frame with a variable scale Gaussian as formulated in Eq. (5)

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (5)$$

Where  $G(x, y, \sigma)$  is the variable scale Gaussian as formulated in Eq. (6)

$$G(x, y, \sigma) = \frac{1}{2\Pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (6)$$

In the next stage, the difference of two scales separated by a constant multiplicative factor  $k$  is computed according to Eq. (7)

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (7)$$

To find the local maxima and minima of  $D(x, y, \sigma)$ , each spot is compared to its eight neighbors in the frame under observation and nine neighbors in the scale above and below it. It is chosen only if it is larger than all of these neighbors or smaller than all of them. An important aspect of SIFT approach is that it produces large numbers of features that densely cover the video frame over the full range of scales and locations [60], [61]. The quantity of features is particularly important edge detection, where the ability to consider objects exhibiting different actions is very crucial.

We extract both spatial and temporal features from a set of testing videos. We then adopt an MLP feed-forward neural network to learn the behavior of these features instead of considering the values of all the pixels. In fact, these features are exploited to learn different classes of human actions. The motivation for exploring MLP is in its substantial ability, through backpropagation, to resist to noise, and the dexterity to generalize. During the training stage, the weights  $W$  and biases  $b$  are updated so that the actual output  $y$  becomes closer to the desired output  $d$ . For this purpose, a cost function is defined as in Eq. (8).

$$E(W, b) = \frac{1}{2} \sum_{i=1}^{n_i} (d_i - y_i^L)^2 \quad (8)$$

The cost function calculates the squared error between the desired and actual output vectors and the backpropagation is gradient descent on the cost function in Eq. (8).

### 3. Experimental Results

Berkeley dataset [14] is an extension of the BSDS300, where the original 300 images are used for training / validation and 200 fresh images, together with human annotations, are added for testing. Each image was segmented by five different subjects on average. Performance is evaluated by measuring Precision / Recall on detected boundaries and three additional region-based metrics.



Figure 2. Berkeley dataset. Two sample images are depicted in the figure from the dataset.

The neural network has been configured using one input layer, two hidden layers and one output layer. The input layer consists of three neurons, each hidden layer consists of three neurons, and a single neuron is considered in the output layer. The adjustment of the neural network in terms of number of layers and number of neuron does not affect the performance significantly. We use an MLP neural network to learn different edges based on our proposed features [62], [63].

The experimental results are presented in Table 1 in terms of precision, recall and F-score. To further elaborate the effectiveness of our proposed method, we carried out experiments considering only LBP and SIFT features separately and the combination of these two features. From the table it is obvious that SIFT feature out performs LBP features since SIFT features are appearance based features while LBP are the texture features. From the table it is also obvious that the combination of both these features improve the performance of our proposed edge detector.

We also report qualitative comparisons of our method with canny edge detector and results are reported in Figure 3. The first column are the input images while the second column are the edges detected by canny detector while the last column shows the results generated by our proposed method. From the Figure, it is obvious that canny edge detection method's output contain a lot of noise while such noise are suppressed in our proposed method while such noise are suppressed in our proposed method.

### 4. Conclusions

In this paper, we proposed an approach for edge detection using our proposed features and an MLP feed-forward neural network. We demonstrated the capability of our approach in capturing the the dynamics of different classes by extracting these features. These features adopt the MLP neural network to learn different edges. The main

advantage of the proposed method is its simplicity and robustness.

Table 1. Our spatio-temporal model. Our proposed spatio-temporal model shows very good performance considering Berkeley dataset. Best results are reported in bold.

Methods	Precision	Recall	F-Score
P-LBP	0.55	0.15	0.45
SIFT	0.58	0.35	0.50
P-LBP + SIFT (proposed)	<b>0.75</b>	<b>0.41</b>	<b>0.69</b>

**References:**

[1] Kimia, B. B., Li, X., Guo, Y., & Tamrakar, A. (2018). Differential Geometry in Edge Detection: accurate estimation of position, orientation and curvature. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[2] Su, B. B., Gu, M. H., Wang, M. M., & Wang, Z. L. (2018, August). Anisotropic Gaussian kernels edge detection algorithm based on the chromatic difference. In *Tenth International Conference on Digital Image Processing (ICDIP 2018)* (Vol. 10806, p. 108062Q). International Society for Optics and Photonics.

[3] Ren, Hongdou, Shengmei Zhao, and Jozef Gruska. "Edge detection based on single-pixel imaging." *Optics express* 26, no. 5 (2018): 5501-5511.

[4] Bui, T. A., & Lee, P. J. (2018, May). Adaptive edge detection algorithm for multi-focus application. In *2018 International Symposium on Consumer Technologies (ISCT)* (pp. 26-28). IEEE.

[5] Canny, John. "A computational approach to edge detection." *IEEE Transactions on pattern analysis and machine intelligence* 6 (1986): 679-698.

[6] Marr, David, and Ellen Hildreth. "Theory of edge detection." *Proc. R. Soc. Lond. B* 207, no. 1167 (1980): 187-217.

[7] Machuca, Raul, and Alton L. Gilbert. "Finding edges in noisy scenes." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1 (1981): 103-111.

[8] El-Sayed, M. A., & Khafagy, M. A. (2011). Using Renyi's entropy for edge detection in level images. *International Journal of Intelligent Computing and Information Sciences*, 11, 1-10.

[9] El-Sayed, M. A. (2013). Study of Edge Detection Based On 2D Entropy. *International Journal of Computer Science Issues (IJCSI)*, 10(3), 1.

[10] El-Sayed, M. A., Bahgat, S. F., & Abdel-Khalek, S. (2013). Novel approach of edges detection for digital images based on hybrid types of entropy. *Applied Mathematics & Information Sciences*, 7(5), 1809.

[11] Pinho, A. J., & Almeida, L. B. (1995, September). Edge detection filters based on artificial neural networks.

In *International Conference on Image Analysis and Processing* (pp. 159-164). Springer, Berlin, Heidelberg.

[12] Qi, Xianbiao, Rong Xiao, Chun-Guang Li, Yu Qiao, Jun Guo, and Xiaoou Tang. "Pairwise rotation invariant co-occurrence local binary pattern." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, no. 11: 2199-2213, 2014.

[13] Lowe, David G. "Object recognition from local scale-invariant features." In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, vol. 2, pp. 1150-1157. Ieee, 1999.

[14] Arbelaez, Pablo, Michael Maire, Charles Fowlkes, and Jitendra Malik. "Contour detection and hierarchical image segmentation." *IEEE transactions on pattern analysis and machine intelligence* 33, no. 5 (2011): 898-916.

[15] Ullah, H., Altamimi, A. B., Uzair, M., & Ullah, M. (2018). Anomalous entities detection and localization in pedestrian flows. *Neurocomputing*, 290, 74-86.

[16] Ullah, H., Ullah, M., & Uzair, M. (2018). A hybrid social influence model for pedestrian motion segmentation. *Neural Computing and Applications*, 1-17.

[17] Ahmad, F., Khan, A., Islam, I. U., Uzair, M., & Ullah, H. (2017). Illumination normalization using independent component analysis and filtering. *The Imaging Science Journal*, 65(5), 308-313.

[18] Ullah, H., Uzair, M., Ullah, M., Khan, A., Ahmad, A., & Khan, W. (2017). Density independent hydrodynamics model for crowd coherency detection. *Neurocomputing*, 242, 28-39.

[19] Khan, S. D., Tayyab, M., Amin, M. K., Nour, A., Basalamah, A., Basalamah, S., & Khan, S. A. (2017). Towards a Crowd Analytic Framework For Crowd Management in Majid-al-Haram. *arXiv preprint arXiv:1709.05952*.

[20] Saqib, M., Khan, S. D., Sharma, N., & Blumenstein, M. (2017, December). Extracting descriptive motion information from crowd scenes. In *2017 International Conference on Image and Vision Computing New Zealand (IVCNZ)* (pp. 1-6). IEEE.

[21] Ullah, M., Ullah, H., Conci, N., & De Natale, F. G. (2016, September). Crowd behavior identification. In *Image Processing (ICIP), 2016 IEEE International Conference on* (pp. 1195-1199). IEEE.

[22] Khan, S. (2016). *Automatic Detection and Computer Vision Analysis of Flow Dynamics and Social Groups in Pedestrian Crowds*.

[22] Arif, M., Daud, S., & Basalamah, S. (2013). Counting of people in the extremely dense crowd using genetic algorithm and blobs counting. *IAES International Journal of Artificial Intelligence*, 2(2), 51.

[23] Ullah, H., Ullah, M., Afridi, H., Conci, N., & De Natale, F. G. (2015, September). Traffic accident detection through a hydrodynamic lens. In *Image Processing (ICIP), 2015 IEEE International Conference on* (pp. 2470-2474). IEEE.

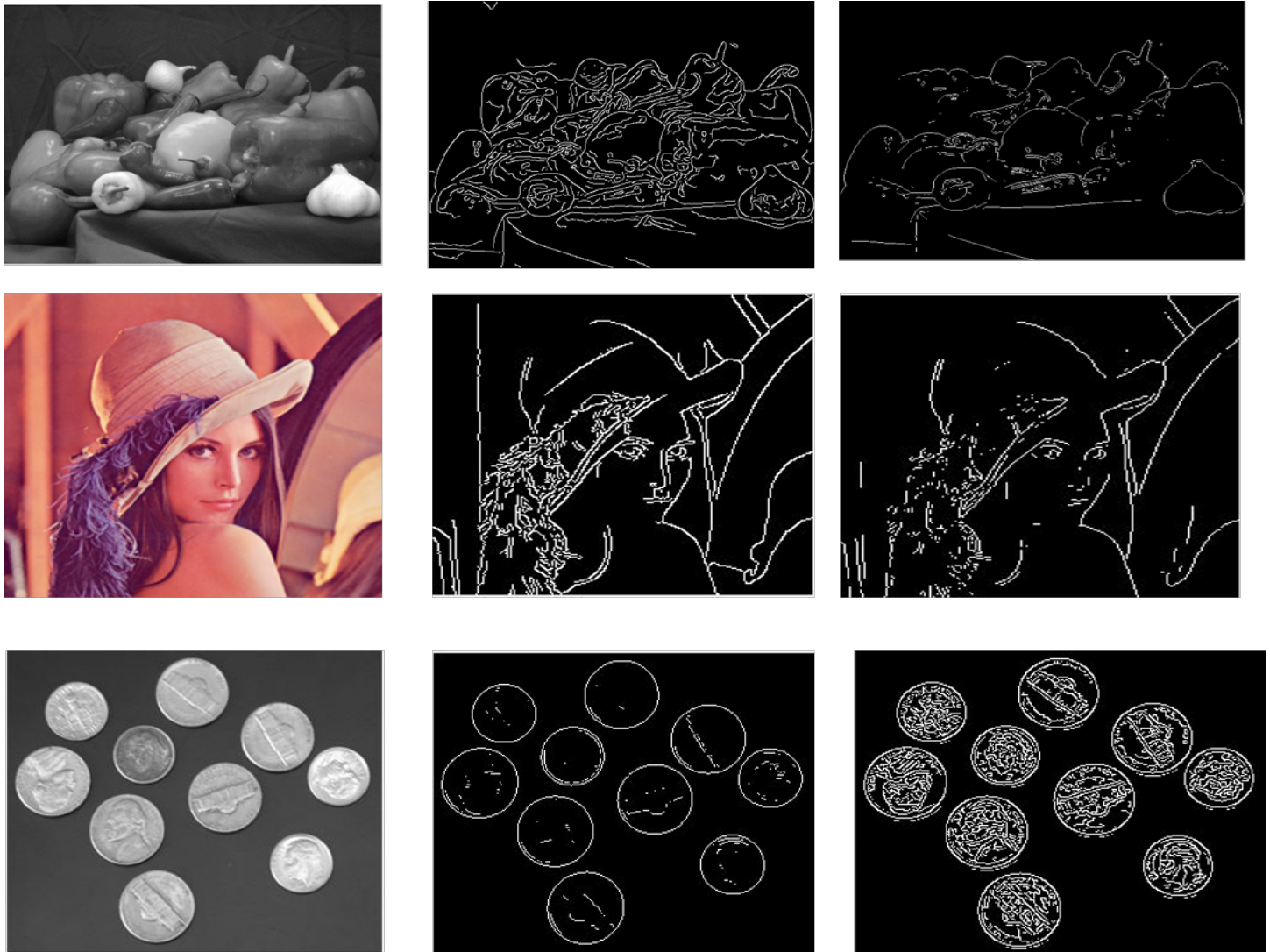
[24] Ullah, H. (2015). *Crowd Motion Analysis: Segmentation, Anomaly Detection, and Behavior Classification* (Doctoral dissertation, University of Trento).

[25] Khan, S. D., Bandini, S., Basalamah, S., & Vizzari, G. (2016). Analyzing crowd behavior in naturalistic conditions: Identifying sources and sinks and characterizing main flows. *Neurocomputing*, 177, 543-563.



- [26] Shimura, K., Khan, S. D., Bandini, S., & Nishinari, K. (2016). Simulation and Evaluation of Spiral Movement of Pedestrians: Towards the Tawaf Simulator. *Journal of Cellular Automata*, 11(4).
- [27] Khan, S. D., Vizzari, G., & Bandini, S. (2016). A Computer Vision Tool Set for Innovative Elder Pedestrians Aware Crowd Management Support Systems. In *AI\* AAL@ AI\* IA* (pp. 75-91).
- [28] Saqib, M., Khan, S. D., Sharma, N., & Blumenstein, M. (2017, August). A study on detecting drones using deep convolutional neural networks. In *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*(pp. 1-5). IEEE.
- [29] Khan, S. D., Vizzari, G., Bandini, S., & Basalamah, S. (2015, October). Detection of social groups in pedestrian crowds using computer vision. In *International Conference on Advanced Concepts for Intelligent Vision Systems* (pp. 249-260). Springer, Cham.
- [30] Khan, S. D., Porta, F., Vizzari, G., & Bandini, S. (2014, September). Estimating Speeds of Pedestrians in Real-World Using Computer Vision. In *International Conference on Cellular Automata* (pp. 526-535). Springer, Cham.
- [31] Khan, S. D., Crociani, L., & Vizzari, G. (2013). Integrated Analysis and Synthesis of Pedestrian Dynamics: First Results in a Real World Case Study. *From Objects to Agents*.
- [32] Khan, S. D., Crociani, L., & Vizzari, G. Pedestrian and Crowd Studies: Towards the Integrated Analysis and Synthesis.
- [33] Ullah, H., Ullah, M., & Conci, N. (2014, September). Dominant motion analysis in regular and irregular crowd scenes. In *International Workshop on Human Behavior Understanding*(pp. 62-72). Springer, Cham.
- [34] Saqib, M., Khan, S. D., & Blumenstein, M. (2017, February). Detecting dominant motion patterns in crowds of pedestrians. In *Eighth International Conference on Graphic and Image Processing (ICGIP 2016)* (Vol. 10225, p. 102251L). International Society for Optics and Photonics.
- [35] Ullah, H., Ullah, M., & Conci, N. (2014, March). Real-time anomaly detection in dense crowded scenes. In *Video Surveillance and Transportation Imaging Applications 2014*(Vol. 9026, p. 902608). International Society for Optics and Photonics.
- [36] Ullah, H., Tenuti, L., & Conci, N. (2013, March). Gaussian mixtures for anomaly detection in crowded scenes. In *Video Surveillance and Transportation Imaging Applications* (Vol. 8663, p. 866303). International Society for Optics and Photonics.
- [37] Rota, P., Ullah, H., Conci, N., Sebe, N., & De Natale, F. G. (2013, September). Particles cross-influence for entity grouping. In *Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21st European* (pp. 1-5). IEEE.
- [38] Ullah, H., & Conci, N. (2013, September). Structured learning for crowd motion segmentation. In *Image Processing (ICIP), 2013 20th IEEE International Conference on* (pp. 824-828). IEEE.
- [39] Ullah, H., & Conci, N. (2012, November). Crowd motion segmentation and anomaly detection via multi-label optimization. In *ICPR workshop on Pattern Recognition and Crowd Analysis*.
- [40] Khan, W., & Ullah, H. (2010). Authentication and Secure Communication in GSM, GPRS, and UMTS Using Asymmetric Cryptography. *International Journal of Computer Science Issues* (IJCSI), 7(3), 10.
- [41] Ullah, H., Ullah, M., Uzair, M., & Rehman, F. (2010). Comparative study: The evaluation of shadow detection methods. *International Journal Of Video & Image Processing And Network Security (IJVIPNS)*, 10(2), 1-7.
- [42] Khan, W., & Ullah, H. (2010). Scientific Reasoning: A Solution to the Problem of Induction. *International Journal of Basic & Applied Sciences*, 10(3), 58-62.
- [43] Uzair, M., Khan, W., Ullah, H., & Rehman, F. U. (2009, December). Background modeling using corner features: An effective approach. In *Multitopic Conference, 2009. INMIC 2009. IEEE 13th International* (pp. 1-5). IEEE.
- [44] Ullah, M., Ullah, H., & Alseadoon, I. M. HUMAN ACTION RECOGNITION IN VIDEOS USING STABLE FEATURES.
- [45] Khan, W., Ullah, H., & Hussain, R. (2013). Energy-Efficient Mutual Authentication Protocol for Handheld Devices Based on Public Key Cryptography. *International Journal of Computer Theory and Engineering*, 5(5), 754.
- [46] Saqib, M., Khan, S. D., & Blumenstein, M. (2016, November). Texture-based feature mining for crowd density estimation: A study. In *Image and Vision Computing New Zealand (IVCNZ), 2016 International Conference on* (pp. 1-6). IEEE.
- [47] Khan, A. (2015). *Coordinated Unmanned Aerial Vehicles for Surveillance of Targets* (Doctoral dissertation, Queen Mary University of London).
- [48] Khan, A., Rinner, B., & Cavallaro, A. (2015, September). Multiscale observation of multiple moving targets using micro aerial vehicles. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on* (pp. 4642-4649). IEEE.
- [49] Ahmad, K., Mekhalfi, M. L., Conci, N., Melgani, F., & Natale, F. D. (2018). Ensemble of Deep Models for Event Recognition. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 14(2), 51.
- [50] Ahmad, K., Conci, N., & De Natale, F. G. (2018). A saliency-based approach to event recognition. *Signal Processing: Image Communication*, 60, 42-51.
- [51] Arif, Muhammad, Sultan Daud, and Saleh Basalamah. "People counting in extremely dense crowd using blob size optimization." *Life Science Journal* 9, no. 3 (2012): 1663-1673.
- [52] Saqib, Muhammad, S. D. Khan, and S. M. Basalamah. "Vehicle Speed Estimation using Wireless Sensor Network." In *INFOCOMP 2011 First International Conference on Advanced Communications and Computation, IARIA*. 2011.
- [53] Khan, Sultan Daud. "Estimating Speeds and Directions of Pedestrians in Real-Time Videos: A solution to Road-Safety Problem." In *CEUR Workshop Proceedings*, p. 1122. 2014.
- [54] Khan, Sultan Daud, and Hyunchul Shin. "Effective memory access optimization by memory delay modeling, memory allocation, and buffer allocation." In *SoC Design Conference (ISOC), 2009 International*, pp. 153-156. IEEE, 2009.
- [55] Khan, Sultan Daud, Giuseppe Vizzari, and Stefania Bandini. "Facing Needs and Requirements of Crowd Modelling:

- Towards a Dedicated Computer Vision Toolset." In *Traffic and Granular Flow'15*, pp. 377-384. Springer, Cham, 2016
- [56] Ullah, Mohib, and Faouzi Alaya Cheikh. "Deep Feature Based End-to-End Transportation Network for Multi-Target Tracking." In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 3738-3742. IEEE, 2018.
- [57] Ullah, Mohib, Mohammed Ahmed Kedir, and Faouzi Alaya Cheikh. "Hand-Crafted vs Deep Features: A Quantitative Study of Pedestrian Appearance Model." In *2018 Colour and Visual Computing Symposium (CVCS)*, pp. 1-6. IEEE, 2018.
- [58] Ullah, Mohib, Ahmed Mohammed, and Faouzi Alaya Cheikh. "PedNet: A Spatio-Temporal Deep Convolutional Neural Network for Pedestrian Segmentation." *Journal of Imaging* 4, no. 9 (2018): 107.
- [59] Ullah, Mohib, and Faouzi Alaya Cheikh. "A Directed Sparse Graphical Model for Multi-Target Tracking." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1816-1823. 2018.
- [60] Ullah, Mohib, Ahmed Kedir Mohammed, Faouzi Alaya Cheikh, and Zhaohui Wang. "A hierarchical feature model for multi-target tracking." In *Image Processing (ICIP), 2017 IEEE International Conference on*, pp. 2612-2616. IEEE, 2017.
- [61] Ullah, Mohib, Faouzi Alaya Cheikh, and Ali Shariq Imran. "Hog based real-time multi-target tracking in bayesian framework." In *2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 416-422. IEEE, 2016.
- [62] Khan, S., Giuseppe Vizzari, Stefania Bandini, and Saleh Basalamah. "Detecting dominant motion flows and people counting in high density crowds." *Journal of WSCG* 22, no. 1 (2014): 21-30.
- [63] Saqib, Muhammad, Sultan Daud Khan, Nabin Sharma, and Michael Blumenstein. "Person Head Detection in Multiple Scales Using Deep Convolutional Neural Networks." In *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-7. IEEE, 2018.



a

b

c

Figure 3: (a) sample input frames (b) output of canny edge detection (c) output of our proposed method.