
ISSN 2047-3338

# Movement of Interest Detection in Crowd Scenes

Abdulrahman Alreshidi[1] and Hina Afridi[2]

[1]College of Computer Science and Engineering, University of Ha'il, Ha'il, Saudi Arabia

[2]COMSATS University, Wah Cantt, Pakistan

[1]ab.alreshidi@uoh.edu.sa, [2]hinaafridi123@yahoo.com

***Abstract*– We introduce a novel method to detect movement of interest in crowd scenes. For this purpose, we consider regions of interest and discretize them into a number of patterns. Furthermore, we investigate a representative movement of key patterns to detect movement of interest. The investigation of region of interest and key patterns help in detecting movement of interest in both regular and irregular scenes. The experimental analysis on a standard dataset reveals the effectiveness of our method.**

***Index Terms*– Region of Interest, Key Patterns, Phase Correlation and Bayes Theorem**

## I. INTRODUCTION

THE crowded gatherings present challenges of greater importance to public safety [1], [2], [3]. These situations are threatening in presence of motion patterns in terms of riots and chaotic acts of crowds in different places [4], [5]. It is significantly important to detect the occurrence of such patterns as early as possible since it can reduce the potential dangerous consequences and can also alert a human operator for monitoring the ongoing situation more effectively. However, the consistent monitoring of video streaming generally requires human operators to watch the displays, which often leads to fatigue, inattention, and failure to identify the occurrence of unwanted motion patterns [6], [7]. On the other hand, significant challenges arise with the substantial amount of surveillance video data, which are difficult and time-consuming for manual analysis [8], [9], [10]. After studying these challenges, an automated movement of interest detection system would fix the aforementioned problems more efficiently.

Rota et al. [25] used influence of particles to detect different groups, whereas Ullah et al. [11], [12] employed a unified model to detect different motion patterns. However, general patterns detection presents difficulty due to the unclear definition of important movements in practical conditions. One may think that the presence of a moving vehicle on a pedestrian pathway is normal, but others may consider it as abnormal motion pattern. In fact, the movement of interest is an observation that does not maintain consistency with other observations over time. Considering the inconsistency between normal and abnormal patterns, one can design normal patterns in an unsupervised or semi-supervised manner, and the pattern that deviates from the model is considered as abnormality. Significant research has been carried out on feature modeling for different patterns. In rural areas where classic target tracking can be well achieved, high-level features, such as corner features, can be used for movement of interest detection. However, in urban areas where not all targets can be accurately tracked, low-level features including histogram of oriented optical flow and gradients, social force models [3], [13], and motion scale-invariant feature transform are robust for extraction, are often used to detect these patterns in videos.

## II. LITERATURE REVIEW

In [14], [15] crowd scene is considered as a dynamic flow field, which is the popular approach recently. Saqib et al. [16] and others [17], [18], [19] focused on the extraction of crowd attributes from the vector fields to describe crowd density, moving directions, and boundaries. In recent years, more attention has shifted towards application-oriented techniques to improve crowd pattern interpretation [20], [21], [22]. In 2014, Ullah et al. [23], [24], [25] first introduced a crowd scene model based on optical flow field, an extension of the flow-filed model for segmenting extremely dense crowd scenes recorded in videos. This model has also been applied in group tracking that contain multiple or intersected crowd entities [26]. Khan et al. [28] off-line dominating crowd moving direction learning algorithm [27] has also been proven an effective flow-based tracking approach. Crowd Kanade-Lucas-Tomasi corners, multi-label optimization and Lagrangian modeled anomaly crowd visual features from flow field information. Those methods demonstrated their potentials in tracking the dynamic crowd under extremely crowded and partial occluded conditions but are bound to pre-defined crowd patterns.

Some methods define motion patterns in videos as Spatio-temporal Volume (STV), which combines global video dynamics into a three-dimensional feature space. For example, [29], [30] introduced a motion labelling method based on the co-occurrence of features. The model is designed as a potential function in the Markov Random Field (MRF) process for anomaly detection. The method in [31], [32] introduced STV-based motion patterns in volumetric environment to represent the spatial-temporal statistical characteristics of pedestrians in crowded scenes. The methods [33], [34], [35], [36] developed a STV-based anomaly location detection approach through using localized cuboids in an unsupervised learning model. These methods are valid approaches when considering spatiotemporal features without parameter settings. Some researchers [37], [38], [41], [42] consider deep learning-based methods for event detection.

## III. METHODOLOGY

The work [39] is motivated by the fact that a sequence of regions of interest (ROI) containing different foreground moving objects is crucial for the detection of significant movement in the crowd. According to the Feng et al. [39] model, we discretize the space movement into a number of patterns which are fitted to ROI for the detection. Each individual movement is modeled as a stochastic concatenation of an arbitrary number of movelet codewords, each one of which may generate significant information. Additionally, Hidden Markov models are used to represent actions and the images they generate. The Viterbi algorithm is considered to investigate the sequence of codewords and movement that are associated to a given observed sequence.

To recognize the observed movements, we observe that the images of the moving crowd are segmented as the foreground in image sequence, which sometimes can be achieved either using background subtraction or independent moving crowd segmentation. For each pair of frames, an observed configuration is defined as the foreground pixels X = {X; Y}, where X and Y accordingly represent the 2D positions of the foreground pixels in the first and second images. In fact, the observed configuration consists of foreground pixels in their entirety and top-down approach is implemented. Given a codeword with shape S and deformed shape S0, we wish to calculate the likelihood of observing X and Y.

It is ensured that any two neighboring parts in crowd have different structures and the two parts having the same structures are located far away from each other. Therefore, each movement can be segmented by its structure and location difference from the others. A movelet is formed by stacking the movements of all crowd visible on a pair of consecutive frames together.

To enhance our method, we also consider the technique proposed by Ogale et al. [40]. Given a sequence, the issue at hand is how to find a representative movement of key patterns to describe the crowd as a whole. For a given video after background removal, we define a keyframe to be a frame where the average of the optical flow magnitude of foreground pixels reaches an extremum. In fact, the optical flow is measured in the reference frame of the foreground, i.e. the mean optical flow of the foreground is first subtracted from the flow value at each foreground pixel. Using the given frames and the 2D optical flow for each frame, we investigate extrema of the discrete function (eq. 1):

$$K_i = \frac{1}{N_i} \sum_{(x,y)\in foreground_i} |\vec{u}_i(x,y) - \vec{u}_i^{mean}| \quad (1)$$

In the equation 1, N is the number of foreground pixels and u is the mean foreground flow in frame. Frames where this value reaches a minimum show flow reversal which occur when the crowd reaches an extreme congestion. Frames at the maxima are points where the crowd is exactly in between two extreme configurations, and hence is undergoing large overall movement.

Given a crowd video sequence consisting of different movement patterns, we perform keyframe extraction on it to obtain an observed sequence of diverse patterns. We then compare each observed pattern with the pattern corresponding to every state. We register pattern sk to pattern si using the phase correlation process to remove 2D translation, rotation and scaling. We also calculate a matching measure between the two patterns, which finds the ratio of their area of intersection to the area of their union. We calculate the probability of a pattern as formulated in eq. 2:

$$P(p_i^v|s_k) = \frac{m(s_k, s_i^v)}{\sum_{all\ s_i^v} m(s_k, s_i^v)} \quad (2)$$

We calculate the observation likelihood in term of Bayes theorem as formulated below in eq. 3.

$$P(s_k|p_i^v) = \frac{P(p_i^v|s_k)P(s_k)}{P(p_i^v)} \quad (3)$$

## IV. EXPERIMENTAL ANALYSIS

We evaluate the performance of our proposed method for movement of interest detection on UCSD benchmark dataset available publicly. The UCSD dataset consists of two subsets: ped1 and ped2. Both subsets represent surveillance videos captured by a fixed camera overlooking pedestrian walkways. In Ped1, people are moving towards and away from the camera, with some perspective distortion and ped2 contains video of people moving parallel to the camera. The resolutions of Ped1 and Ped2 are 158x238 and 240x360, respectively. The normal motion patterns appearing in the dataset is sequences of pedestrians on the walkways, with a varying density from sparse to very dense. The non-pedestrian entities include cyclists, skaters, vehicles, people walking on a lawn. The appearance of all non-pedestrian entities occurs naturally, i.e., they were not staged or synthesized for data set collection. The video footage of each scene is divided into clips of 120-200 frames. Ped1 consists of 34 training video clips and 36 testing video clips; whereas ped2 contains 16 training video clips and 12 testing video clips.

Quantitative analysis in term of Equal error rate (EER) for frame-level criterion for our proposed method for both subsets, Ped1 and Ped2, are presented in the Table I.

Table I: Results in terms of EER

| Sub-datasets | Our Method |
|---|---|
| Ped 1 | 48 |
| Ped 2 | 51 |
| Average | 49.5 |

## V. CONCLUSIONS

In this paper, we presented a robust method for movement of interest detection using UCSD dataset. We would like to extend it in the future to improve performance and further evaluate it on other datasets.

## REFERENCES

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

[2]. Khan, W., Ullah, H., Ahmad, A., Sultan, K., Alzahrani, A. J., Khan, S. D., & Abdulaziz, S. (2018). CrashSafe: a formal model for proving crash-safety of Android applications. Human-centric Computing and Information Sciences, 8(1), 21.

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

[4]. 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.

[5]. 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.

[6]. 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.

[7]. 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.

[8]. 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.

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

[10]. 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.

[11]. 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.

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

[13]. 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.

[14]. 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).

[15]. 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).

[16]. 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.

[17]. 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.

[18]. 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.

[19]. 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.

[20]. Khan, S. D., Crociani, L., & Vizzari, G. Pedestrian and Crowd Studies: Towards the Integration of Automated Analysis and Synthesis.

[21]. 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.

[22]. 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.

[23]. 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.

[24]. 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.

[25]. 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.

[26]. 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.

[27]. 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.

[28]. 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.

[29]. 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.

[30]. Khan, W., & Ullah, H. (2010). Scientific Reasoning: A Solution to the Problem of Induction. International Journal of Basic & Applied Sciences, 10(3), 58-62.

[31]. 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.

[32]. Ullah, M., Ullah, H., & Alseadoon, I. M. HUMAN ACTION RECOGNITION IN VIDEOS USING STABLE FEATURES.

[33]. Khan, W., Ullah, H., & Hussain, R. (2013). Energy-Efficient Mutual Authentication Protocol for Handhled Devices Based on Public Key Cryptography. International Journal of Computer Theory and Engineering, 5(5), 754.

[34]. 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.

[35]. Ullah, M., & Cheikh, F. A. (2018). 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).

[36]. Ullah, M., Mohammed, A. K., Cheikh, F. A., & Wang, Z. (2017, September). A hierarchical feature model for multi-target tracking. In Image Processing (ICIP), 2017 IEEE International Conference on (pp. 2612-2616). IEEE.

[37]. Ullah, M., Mohammed, A., & Alaya Cheikh, F. (2018). PedNet: A Spatio-Temporal Deep Convolutional Neural Network for Pedestrian Segmentation. Journal of Imaging, 4(9), 107.

[38]. Ullah, M., & Cheikh, F. A. (2018, October). 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.

[39]. Feng, X., & Perona, P. (2002, June). Human action recognition by sequence of movelet codewords. In null (p. 717). IEEE.

[40]. Ogale, A., Karapurkar, A., Guerra-Filho, G., & Aloimonos, Y. (2004, November). View-invariant identification of pose sequences for action recognition. In Video Analysis and Content Extraction Workshop (VACE) (Vol. 14).

[41]. Ullah, M., Cheikh, F. A., & Imran, A. S. (2016, August). 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.

[42]. Nadeem, M., Aqil, B., & Kamran, A. (2010). A Logical Temporal Relational Data Model. IJCSI International Journal of Computer Science Issues, 7(1).

**Abdulrahman Alreshidi** did his PhD from King's College London, UK. Currently, he is working as an Assistant Professor in the college of computer science and engineering at the University of Ha'il, Saudi Arabia. He is also working as the Vice-Dean of Quality Assurance in the same college. He is the author of many peer reviewed journals.

**Hina Afridi** did her MSc in Telecommunication Engineering from the University of Trento, Italy. She is working as a researcher in COMSATS University, Wah Cantt, Pakistan.