American Journal of Engineering Research (AJER) ***2018***
American Journal of Engineering Research (AJER)
e-ISSN: 2320-0847 p-ISSN : 2320-0936
Volume-7, Issue-11, pp-114-118
www.ajer.org
Research Paper Open Access

Crowd behavior identification in the wild

Abdullah J. Alzahrani[1], Jasim Khan[2]
[1]*College of Computer Science andEngineering*
University of Ha'il, Saudi Arabia
[2]*Siemens S.P.A, Milano-Partita IVA n. 00751160151,*
Informazioni Corporate, Italy
Corresponding Author:Abdullah J. Alzahrani

ABSTRACT:*In this paper we present a novel method to recognize crowd behaviors using Fisher kernel combine with other spatio-temporal features. The method exploits these features to extract useful information uniformly distributed on the video frame. The temporal features represent the rendering of trajectories traveled by the individuals and the spatial features represent the density of neighboring individuals in the predefined proximity. This formulation is used to model the behavior of the crowd. The feature extraction process is computationally affordable, thus suitable to be applied in real-time applications for behavior analysis in crowded scenes. The experimental evaluation is conducted on a set of benchmark video sequences commonly used for crowd behavior identification. We show that our approach presents significant performance considering these video sequences.*
KEYWORDS: *Crowd, motion analysis, behavior identification, spatio-temporal features.*

Date Of Submission:02-11-2018 Date Of Acceptance: 16-11-2018

I. INTRODUCTION

More than half of the people in the world live in densely populated areas [1][2][3][4]. Hence, automated detection of anomalous events generated by self-organization phenomena resulting from the interactions of many individuals, can cause significant hindrance in the flow [4][5][6]. This makes necessary to provide more vigilant surveillance, possibly in lieu of, or as an assistance to, human operators. However, there is a lack of empirical studies of crowded scenes where besides basic motion segmentation, also the analysis of more structured behaviors, such as the formation of lanes, or the detection of oscillations at bottlenecks, is decisive for the safety of people during, for example, the access to or exit from mass events, or in situations of emergency evacuation [7][8][9][10]. Congested conditions can possibly trigger crowd disasters arising from the maximum density and irregular flow of crowd [11][12]. Moreover, the behavior of the crowd may transition from one state of collective behavior to a qualitatively different behavior depending on the density of crowd. Such transitions typically occur when individuals in the crowd accumulate, propagate, or uniformly move with the flow [13][14][15]. Activity analysis and scene understanding entail object detection, tracking and activity recognition [16][17][18]. These approaches, requiring low-level motion features, appearance features, or object trajectories, render good performance in low density crowd scenes, but fail in real-world high density crowd scenes. Some recent works [19][20][21] presents an offline data-driven approach for crowd videos to learn crowd motion patterns by performing long-term analysis[22][23]. The approach tracks individuals in dense crowded scenes, showing typical and rare behaviors. Other related works [24][25][26] proposed an interdisciplinary framework for the analysis of the crowd, which integrates benefits of simulation techniques, pedestrian detection and tracking, dense crowd detection and event detection [27][28].

Considering the difficulty in performing detection and tracking in crowded environments [29][30], the research has focused on gathering the motion information at a higher scale, thus not associating it to single objects, but considering the crowd as a single entity [31][32]. These approaches often require low-level features such as multi-resolution histograms [33][34], spatio-temporal volumes [35][36], and appearance or motion descriptors [37][38][39]. In the work [40], the optical flow constraint is exploited to estimate a conditional probability of the spatio-temporal intensity change [41]. Furthermore, motion estimation and segmentation are integrated into a functional minimization strategy based on a Bayesian framework. In [42][43], authors used a

mixture of dynamic textures to fit a video sequence and then assigned homogeneous motion regions to the mixture components [44]. However, the methods presented in [45][46] are only targeted at addressing the cases of simple motion patterns. In [47][48], motion segmentation is performed without relying on the optical flow. In [48], a dynamic texture model is used to measure the similarity between neighboring spatio-temporal patches. These patches are grouped by connected component analysis, resulting into over segmentation in presence of low density crowd. In [53], the authors proposed a method to perform multi-target tracking in crowd using time integration of the dynamical system defined by the optical flow.

II. PROPOSED FRAMEWORK

The Fisher kernel/vector (FV) [49] is a function that measures the similarity of two patches or objects on the basis of sets of measurements for each patch or object. In this process, the class for a new object can be computed by minimizing, across classes, an average of the Fisher kernel distance from the new patch to each known member of the given class. In fact, the FV is an object or patch representation obtained by pooling local image features. It is frequently both as a local and global image descriptor in visual classification.

While the FV can be extracted as a special, approximate, and improved case of the general Fisher Kernel framework, it is easy to demonstrate directly. Let I be a set of different dimensional feature vectors (e.g. SURF descriptors) extracted from an image. Let Θ be the parameters of a Gaussian Mixture Model fitting the distribution of descriptors. The GMM connects each vector xi to a mode k in the mixture with a strength given by the posterior probability:

$$q_{ik} = \frac{\exp\left[-\frac{1}{2}(\mathbf{x}_i - \mu_k)^T \Sigma_k^{-1}(\mathbf{x}_i - \mu_k)\right]}{\sum_{t=1}^{K} \exp\left[-\frac{1}{2}(\mathbf{x}_i - \mu_t)^T \Sigma_k^{-1}(\mathbf{x}_i - \mu_t)\right]}.$$

For each mode k, consider the mean and covariance deviation vectors as formulated below:

$$u_{jk} = \frac{1}{N\sqrt{\pi_k}} \sum_{i=1}^{N} q_{ik} \frac{x_{ji} - \mu_{jk}}{\sigma_{jk}},$$

$$v_{jk} = \frac{1}{N\sqrt{2\pi_k}} \sum_{i=1}^{N} q_{ik} \left[\left(\frac{x_{ji} - \mu_{jk}}{\sigma_{jk}} \right)^2 - 1 \right]$$

where j=1,2,…,D spans the vector dimensions. The FV of the image I is the pooling of the vectors uk and then of the vectors vk for each of the K in the Gaussian mixtures which is formulated below:

$$\Phi(I) = \begin{bmatrix} \vdots \\ \mathbf{u}_k \\ \vdots \\ \mathbf{v}_k \\ \vdots \end{bmatrix}$$

To detect spatio-temporal events, Laptav [50] built on the idea of the Harris and Forstner interest point operators [51][52] and detect local structures in space-time where the object or patch values have local variations in both space and time. The spatio-temporal extents of the detected events are computed and their scale-invariant spatio-temporal descriptors are calculated. Using such descriptors, events are classified and video representation is modeled in terms of labeled space-time points. For the problem of behavior detection, we illustrate how this method allows for detection of these behaviors in different scenes with occlusions and dynamic backgrounds.

The motivation of the Harris interest point detector is to find locations in a spatial image where the image values have high magnitude variations in both directions. For a given scale of observation, such interest

points can be computed from a windowed second moment matrix integrated at a specific scale as formulated below:

$$\mu^{sp} = g^{sp}(\cdot;\, \sigma_i^2) * \begin{pmatrix} (L_x^{sp})^2 & L_x^{sp} L_y^{sp} \\ L_x^{sp} L_y^{sp} & (L_y^{sp})^2 \end{pmatrix}$$

The L's are Gaussian derivatives as formulated below:

$$L_x^{sp}(\cdot;\, \sigma_l^2) = \partial_x (g^{sp}(\cdot;\, \sigma_l^2) * f^{sp})$$
$$L_y^{sp}(\cdot;\, \sigma_l^2) = \partial_y (g^{sp}(\cdot;\, \sigma_l^2) * f^{sp})$$

In the equation, g is the spatial Gaussian kernel as formulated below:

$$g^{sp}(x, y;\, \sigma^2) = \frac{1}{2\pi\sigma^2} \exp(-(x^2 + y^2)/2\sigma^2)$$

The concept of interest points in the spatial domain can be prolonged into the spatio-temporal domain by requiring the object or patch values in space-time to have greater variations in both the spatial and the temporal dimensions. Points with such characteristics will be spatial interest points with a unique location in time corresponding to the moments with non-constant motion of the image in a local spatio-temporal neighborhood.

III. EXPERIMENTS

To validate the performance of our approach, we have conducted the experiments on a set of 22 crowd video sequences extracted from benchmark datasets, commonly used for crowd analysis, including videos from PETS2009, UCSD, and UCF. We have also conducted experiments on video sequences from the UCD dataset, a collection of video sequences acquired in the university campus. The UCD dataset contains video sequences representing flows of students moving outdoor across two buildings. Furthermore, we also evaluated the performance on two real-world video sequences and two synthetic video sequences downloaded from Youtube in order to validate the generalization property of the proposed approach.

For the extraction of the spatio-temporal features for each particle, the resolution of the pixels in original images is not changed. Additionally, each video sequence is partitioned into segments of fixed-length, set at 160 frames (about 6 seconds depending on the frame rate). However, movement of individuals in 6 video sequence is very swift, representing strong transitions between consecutive frames. Therefore, these video sequences are partitioned into segments of 60 frames each instead. There numbers are determined empirically.

For each particle, a two-dimensional Gaussian filter, with variance 1 and size 11 x 11, is applied to reduce noise and engender a consistent density map at the end of particle advection. To extract the features, normalization is applied and the centroid of the image is recognized as a peak. The analysis of the extracted peak shows lane behaviors as the crowd follows a straight path. In the ring/arch behavior, crowd exhibits ring/arch behavior since individuals move in the curved direction. Similarly, in the bottleneck behavior, all the individuals/vehicles converge to a single location.

To evaluate the performance of our approach, presented the results of our methods in the Table below. As can be seen, out method present very good results for crowd behavior identification.

Datasets	Lane behaviors	Ring behaviors	Bottleneck	Our method (Correctly detected behaviors)
PETS2009	25	0	5	Lane = 18, Bottleneck=2
UCSD	65	0	8	L=35, B=3
UCF	15	2	4	L=10, R=1, B=1
UCD	31	1	3	L=10, R=0, B=2
Youtube	13	5	6	L=8, R=2, B=2

IV. CONCLUSION

In this paper we present an approach for crowd behaviors identification using a combination of Fisher kernel and spatio-temporal features. We conducted experiments on a large set of datasets consisting of three important crowd behaviors including lane, ring/arch, and bottleneck. Our method shows very good results after conducted experiments on the video sequences from these datasets. In our future work, we would like to extend our method in order to consider other crowd behaviors including fountain congestion.

REFERENCES

[1]. Rota, Paolo, HabibUllah, Nicola Conci, NicuSebe, and Francesco GB De Natale. "Particles cross-influence for entity grouping." In Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21st European, pp. 1-5. IEEE, 2013.

[2]. Saqib, Muhammad, Sultan Daud Khan, Nabin Sharma, and Michael Blumenstein. "Extracting descriptive motion information from crowd scenes." In 2017 International Conference on Image and Vision Computing New Zealand (IVCNZ), pp. 1-6. IEEE, 2017.

[3]. Shimura, Kenichiro, Sultan Daud Khan, StefaniaBandini, and Katsuhiro Nishinari. "Simulation and Evaluation of Spiral Movement of Pedestrians: Towards the Tawaf Simulator." Journal of Cellular Automata 11, no. 4 (2016).

[4]. Khan, Wilayat, HabibUllah, Aakash Ahmad, Khalid Sultan, Abdullah J. Alzahrani, Sultan Daud Khan, Mohammad Alhumaid, and Sultan Abdulaziz. "CrashSafe: a formal model for proving crash-safety of Android applications." Human-centric Computing and Information Sciences 8, no. 1 (2018): 21.

[5]. Ullah, Habib, MohibUllah, and Muhammad Uzair. "A hybrid social influence model for pedestrian motion segmentation." Neural Computing and Applications (2018): 1-17.

[6]. Ahmad, Fawad, Asif Khan, IhteshamUl Islam, Muhammad Uzair, and HabibUllah. "Illumination normalization using independent component analysis and filtering." The Imaging Science Journal 65, no. 5 (2017): 308-313.

[7]. Zhong, W., Lu, H., & Yang, M. H. (2014). Robust object tracking via sparse collaborative appearance model. IEEE Transactions on Image Processing, 23(5), 2356-2368.

[8]. Ullah, Habib, Muhammad Uzair, MohibUllah, Asif Khan, Ayaz Ahmad, and Wilayat Khan. "Density independent hydrodynamics model for crowd coherency detection." Neurocomputing 242 (2017): 28-39.

[9]. Khan, Sultan Daud, Muhammad Tayyab, Muhammad Khurram Amin, AkramNour, AnasBasalamah, SalehBasalamah, and Sohaib Ahmad Khan. "Towards a Crowd Analytic Framework For Crowd Management in Majid-al-Haram." arXiv preprint arXiv:1709.05952 (2017).

[10]. Ullah, Mohib, HabibUllah, Nicola Conci, and Francesco GB De Natale. "Crowd behavior identification." In Image Processing (ICIP), 2016 IEEE International Conference on, pp. 1195-1199. IEEE, 2016.

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

[12]. Ullah, Habib, Ahmed B. Altamimi, Muhammad Uzair, and MohibUllah. "Anomalous entities detection and localization in pedestrian flows." Neurocomputing 290 (2018): 74-86.

[13]. Arif, Muhammad, Sultan Daud, and SalehBasalamah. "Counting of people in the extremely dense crowd using genetic algorithm and blobs counting." IAES International Journal of Artificial Intelligence 2, no. 2 (2013): 51.

[14]. Ullah, Habib, MohibUllah, HinaAfridi, Nicola Conci, and Francesco GB De Natale. "Traffic accident detection through a hydrodynamic lens." In Image Processing (ICIP), 2015 IEEE International Conference on, pp. 2470-2474. IEEE, 2015.

[15]. Peng, H., Li, B., Ji, R., Hu, W., Xiong, W., & Lang, C. (2013, July). Salient Object Detection via Low-Rank and Structured Sparse Matrix Decomposition. In AAAI (pp. 796-802).

[16]. Ullah, Habib. "Crowd Motion Analysis: Segmentation, Anomaly Detection, and Behavior Classification." PhD diss., University of Trento, 2015.

[17]. Khan, Sultan D., StefaniaBandini, SalehBasalamah, and Giuseppe Vizzari. "Analyzing crowd behavior in naturalistic conditions: Identifying sources and sinks and characterizing main flows." Neurocomputing 177 (2016): 543-563.

[18]. Khan, Sultan Daud, Giuseppe Vizzari, and StefaniaBandini. "A Computer Vision Tool Set for Innovative Elder Pedestrians Aware Crowd Management Support Systems." In AI* AAL@ AI* IA, pp. 75-91. 2016.

[19]. Saqib, Muhammad, Sultan Daud Khan, Nabin Sharma, and Michael Blumenstein. "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, 2017.

[20]. Khan, Sultan Daud, Giuseppe Vizzari, StefaniaBandini, and SalehBasalamah. "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, 2015.

[21]. Khan, Sultan Daud, Fabio Porta, Giuseppe Vizzari, and StefaniaBandini. "Estimating Speeds of Pedestrians in Real-World Using Computer Vision." In International Conference on Cellular Automata, pp. 526-535. Springer, Cham, 2014.

[22]. Khan, Sultan D., Luca Crociani, and Giuseppe Vizzari. "Integrated Analysis and Synthesis of Pedestrian Dynamics: First Results in a Real World Case Study." From Objects to Agents (2013).

[23]. Khan, Sultan D., Luca Crociani, and Giuseppe Vizzari. "PEDESTRIAN AND CROWD STUDIES: TOWARDS THE INTEGRATION OF AUTOMATED ANALYSIS AND SYNTHESIS."

[24]. Ullah, Habib, MohibUllah, and Nicola Conci. "Dominant motion analysis in regular and irregular crowd scenes." In International Workshop on Human Behavior Understanding, pp. 62-72. Springer, Cham, 2014.

[25]. Saqib, Muhammad, Sultan Daud Khan, and Michael Blumenstein. "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, 2017.

[26]. Ullah, Habib, MohibUllah, and Nicola Conci. "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, 2014.

[27]. Ullah, Habib, LorenzaTenuti, and Nicola Conci. "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, 2013.

[28]. Ullah, Habib, and Nicola Conci. "Structured learning for crowd motion segmentation." In Image Processing (ICIP), 2013 20th IEEE International Conference on, pp. 824-828. IEEE, 2013.

[29]. Ullah, Habib, and Nicola Conci. "Crowd motion segmentation and anomaly detection via multi-label optimization." In ICPR workshop on Pattern Recognition and Crowd Analysis. 2012.

[30]. Khan, Wilayat, and HabibUllah. "Authentication and Secure Communication in GSM, GPRS, and UMTS Using Asymmetric Cryptography." International Journal of Computer Science Issues (IJCSI) 7, no. 3 (2010): 10.
[31]. Ullah, Habib, MohibUllah, Muhammad Uzair, and F. Rehman. "Comparative study: The evaluation of shadow detection methods." International Journal Of Video & Image Processing And Network Security (IJVIPNS) 10, no. 2 (2010): 1-7.
[32]. Khan, Wilayat, and HabibUllah. "Scientific Reasoning: A Solution to the Problem of Induction." International Journal of Basic & Applied Sciences 10, no. 3 (2010): 58-62.
[33]. Uzair, Muhammad, Waqas Khan, HabibUllah, and Fasih-ur-Rehman. "Background modeling using corner features: An effective approach." In Multitopic Conference, 2009. INMIC 2009. IEEE 13th International, pp. 1-5. IEEE, 2009.
[34]. Ullah, Mohib, HabibUllah, and Ibrahim M. Alseadoon. "HUMAN ACTION RECOGNITION IN VIDEOS USING STABLE FEATURES."
[35]. Khan, Wilayat, HabibUllah, and RiazHussain. "Energy-Efficient Mutual Authentication Protocol for Handhled Devices Based on Public Key Cryptography." International Journal of Computer Theory and Engineering 5, no. 5 (2013): 754.
[36]. Arif, Muhammad, Sultan Daud, and SalehBasalamah. "People counting in extremely dense crowd using blob size optimization." Life Science Journal 9, no. 3 (2012): 1663-1673.
[37]. 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.
[38]. 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.
[39]. Khan, Sultan Daud, and Hyunchul Shin. "Effective memory access optimization by memory delay modeling, memory allocation, and buffer allocation." In SoC Design Conference (ISOCC), 2009 International, pp. 153-156. IEEE, 2009.
[40]. Khan, Sultan Daud, Giuseppe Vizzari, and StefaniaBandini. "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.
[41]. 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.
[42]. Ullah, Mohib, and FaouziAlayaCheikh. "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.
[43]. Ullah, Mohib, Mohammed Ahmed Kedir, and FaouziAlayaCheikh. "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.
[44]. Ullah, Mohib, Ahmed Mohammed, and FaouziAlayaCheikh. "PedNet: A Spatio-Temporal Deep Convolutional Neural Network for Pedestrian Segmentation." Journal of Imaging 4, no. 9 (2018): 107.
[45]. Ullah, Mohib, and FaouziAlayaCheikh. "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.
[46]. Ullah, Mohib, Ahmed Kedir Mohammed, FaouziAlayaCheikh, 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.
[47]. Ullah, Mohib, FaouziAlayaCheikh, 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.
[48]. Datta, A., Shah, M., & Lobo, N. D. V. (2002). Person-on-person violence detection in video data. In Pattern Recognition, 2002. Proceedings. 16th International Conference on (Vol. 1, pp. 433-438). IEEE.
[49]. F. Perronnin and C. Dance. Fisher kenrels on visual vocabularies for image categorizaton. In Proc. CVPR, 2006.
[50]. Laptev, I. (2005). On space-time interest points. International journal of computer vision, 64(2-3), 107-123.
[51]. W. A. Forstner and E. Gulch. A fast operator for detection and precise location of distinct points, corners and centers of circular features. In ISPRS, 1987.
[52]. C. Harris and M.J. Stephens. A combined corner and edge detector. In Alvey Vision Conference, pages 147–152, 1988.
[53]. Ullah, M., Ullah, H., &Alseadoon, I. M. HUMAN ACTION RECOGNITION IN VIDEOS USING STABLE FEATURES.

Abdullah J. Alzahrani. "Crowd behavior identification in the wild"American Journal of Engineering Research (AJER), vol. 7, no. 11, 2018, pp. 114-118