CHARACTERIZING HUMAN BEHAVIOURS USING STATISTICAL MOTION DESCRIPTOR

Eissa Jaber Alreshidi[1] and Mohammad Bilal[2]

[1]Assistant Professor at College of Computer Science and Engineering, University of Hail, Hail, P.O. Box 2440, Saudi Arabia; [2]Comsats University, Pakistan

ABSTRACT

Identifying human behaviors is a challenging research problem due to the complexity and variation of appearances and postures, the variation of camera settings, and view angles. In this paper, we try to address the problem of human behavior identification by introducing a novel motion descriptor based on statistical features. The method first divide the video into N number of temporal segments. Then for each segment, we compute dense optical flow, which provides instantaneous velocity information for all the pixels. We then compute Histogram of Optical Flow (HOOF) weighted by the norm and quantized into 32 bins. We then compute statistical features from the obtained HOOF forming a descriptor vector of 192-dimensions. We then train a non-linear multi-class SVM that classify different human behaviors with the accuracy of 72.1%. We evaluate our method by using publicly available human action data set. Experimental results shows that our proposed method out performs state of the art methods.

KEYWORDS

Support vector machine, motion descriptor, features, human behaviours

1. INTRODUCTION

Automatic recognition of human actions play the importance role and is the most dominating research topic in computer vision research [1], [2], [3]. It has a wide range of application in automated surveillance [7], [8], [9] [10], human computer interaction [4], [5], [6], and video indexing [11], [12], [13] and video retrieval [14], [15], [16]. Human perform action for a specific purpose. For example, a patient is doing an exercise by interacting with environment by using his/her hands, arms, legs, and other body parts. An action can be either observed with the bare eyes or measured by using camera. With the bared eye, we can easily understand and classify that action into a specific category. For example, a person is walking or running, we can easily discriminate walking behavior from the running behavior.

For the video surveillance and smart rehabilitation, it is important to observe and monitor human actions for a long period [17], [18]. It is humanly impossible to monitor these actions for long duration due to limited human capabilities [19], [20]. Therefore, there is an increased interest to automate this process by using surveillance camera installed in different location of scene.

One of the ultimate goals of artificial intelligence research is to design a virtual machine [21], [22], [23] that can accurately analyze and understand humans' actions, so to reduce to human labor. For example, a patient is undergoing a rehabilitation exercise at home, and a virtual analyst that can understand and recognize his /her behaviors analyzes all his activities. With the help of such virtual analyst, we can prevent the patient from injuries. Such virtual analyst would be greatly beneficial as it saves the trips cost and medical cost. Other important applications including visual surveillance, entertainment, and video retrieval also need to analyze human actions in videos. Action recognition and prediction algorithms have wide range of applications. Several work reported in literature [24], [25], [26], [27] with aim to automatically classify human

actions that have substantially reduce the human labor in analyzing a large-scale of video data and provide understanding on the current state and future state.

Public safety and security is now becoming more important and in place under surveillance [28], certain human actions are not allowed [29]. In order to ensure public safety, surveillance camera are generally mounted on several places around the area under surveillance. With this camera network, action recognition and prediction algorithms may help in capturing malicious activities of human and therefore can reduce the risk caused by criminal actions. Therefore, acknowledging the importance of automatic human behavior analysis, several algorithms are proposed to solve this problem. Cutler et al. [30] [31] detect and recognize the periodic motion in very-low-resolution images. They first compute self-similarity, which evolves in time, and from this analysis, they showed whether an action is periodic. The problem with this method is that only used appearance features while computing similarity of appearance based features cannot discriminate the variation of posture and appearance between objects. Therefore, some of the researches consider the motion gradient information to classify actions. Efros et al. [32] introduced motion descriptor based on the optical flow and motion similarity. The measured noisy optical flow computed among the consecutive frames is smoothed out in four separated channels. Then a spatio-temporal motion descriptors is computed which then can classify using nearest-neighbor. They applied method on low-resolution videos and retrieve the postures of similar actions from action database. In [33], the method uses mid-level motion features, then threshold is used to extract low-level motion features classifier. Evaluation is performed with the dataset [34].. The motion features are similar to motion descriptors in [35]. The features are extracted using a variant of AdaBoost which focuses on the local regions. Other methods extract noisy motion through optical flow by using histograms of orientations. Chaudhry et al. [36] proposed a histogram of oriented Optical flow (HOOF) and used Binet-Cauchy kernels to classify human actions. One of the advantage of HOOF method is that it can alleviate the effect of noise, scale and motion variation. We propose an approach which is similar to HOOF but with significant differences: [37] computes histogram of optical flow (HOOF) from the time series data while our approach compute statistical features from the HOOF [38], [39], [40], which is very different from our approach.

Our main contribution are as follows: Our approach utilize dense optical flow information to build a motion descriptor, which can be used for identifying human behaviors. After computing optical flow, we build a motion descriptor by computing statistics from histogram of optical flow weighted by the norm of the velocity. The resultant statistics are concatenated representing motion descriptor, which are then used as the input of a SVM binary.

This paper is organized as follows: Section 2 introduces the proposed descriptor, section 3 demonstrates the effectiveness of the method and the last section concludes with a discussion and possible future works.

2. PROPOSED METHODOLOGY

The proposed methodology starts by computing a dense optical flow between two consecutive frames using the local jet feature space approach [41]. The advantage of computing dense optical flow is it allows us to segment the region, which contain motion information. After extracting foreground information, motion orientation histogram is then calculated, using typically 32 directions. Every direction bin is weighted by the norm of the flow vector. Finally, we compute list of statistics from HOOF, which will be our final motion descriptor. Later on, we train non-linear classifier that will classify different human action in different categories.

2.1 OPTICAL FLOW COMPUTATION

The first step to extract motion information is the computation of dense optical flow between two consecutive frames. For computing optical flow we employ methods [42] where gray value consistency, gradient constancy and smoothness in multi-scale constraints are used to compute highly accurate optical flow. Consider a feature point i in the frame associated to time t of a segment: its flow vector Zi,t = (Xi,t, Vi,t) includes the location of feature and its velocity is represented by Vx and Vy. Where Vx represents the change in horizontal direction and Vy represents change in vertical direction. After computing optical flow for each pixel, we have now motion field where high magnitude represents the pixels corresponds to foreground and lower magnitude pixels represents the pixels corresponds to the background.

2.2 PARTICLE ADVECTION

After computing optical flow, the next step is to generate dense and long trajectories based on the optical flow [43], [44]. For doing so, we overlay grid of particles over the first optical flow field where each initial location (horizontal and vertical location) of the particle represents the source point. In order to generate dense trajectories, we keep the size of the grid as same as resolution of the frame. The size of particle is same as size of the pixel. This arrangement will incur computational costs. In order to reduce the computational cost and to generate dense trajectories, we reduce the resolution of the grid by dividing the size of the grid by a positive constant. During the advection process, we keep two separate flow maps, one to keep the horizontal coordinates and other map keeps track of vertical components of the trajectory. These map in general store the initial and subsequent positions of the point trajectories evolved during the process of particle advection.

The trajectories obtained through this process are suitable for structured crowds but in the unstructured crowds [45], [46], where people move in different directions, such trajectories do not represent the actual motion flow. The reason is that in unstructured crowds, the people move in arbitrary directions and in most of the case there is chance that particle will lose its path and become the part of different motion pattern moving towards a complete different direction. In this case, the trajectory is unreliable and erroneous. In order to avoid the above problem, we modify the above equation in the following way:

$$X_{(i,t+1)} = X_{(i,t)} + F(X_{(i,t)}) * B_i$$

Trajectories obtained using the above equation will avoid errors caused by particle drifting from one pedestrian flow to different motion pattern. The trajectories obtained through this method are precise, accurate but longer in length.

After particle advection, trajectories obtained corresponds to foreground while some of these trajectories correspond to the background of the scene or noise which are actually are not the part of actual motion pattern [47], [48], [49]. Therefore, in order to refine the obtained set of trajectories, we compute length of each trajectory by calculating the Euclidean distance between the source and sink points of the trajectories. We observed from our experiments that trajectories corresponds to noise and background are generally shorter in length. We exploit this information by setting a threshold value on the length of trajectories and suppress those trajectories whose length are shorter than the specified threshold

2.3 COMPUTING VELOCITY ORIENTATION HISTOGRAM

The next step is to obtain the distribution of orientation of each trajectory, since trajectory captures the spatial-temporal information; therefore, we need to estimate the distribution of orientations that can provide an aid in identifying type of motion. Let for non-zero vector V computed between any two consecutive points of the trajectory. Let $\phi(V)$ denotes the quantized orientation. Similar to HOG descriptor [50], we compute the histogram of optical flow vectors of each trajectory weighted by the vector norm:

$$H_t(\omega) = \frac{\Sigma_{\{x;\, \phi(V_t(x))=\omega\}} \|V_t(x)\|}{\Sigma_{\{x;\, \|V_t(x)\|>0\}} \|V_t(x)\|}$$

Where $\omega \in \{\omega o \dots \omega N-1\}$. Where ωN represents the number of orientations, which is set to 32 in our experiments. We capture the motion information similar to the HOOF descriptor of [51], except that the HOOF descriptor is not symmetrical. In other words, HOOF proposed [52] could not differentiate between the left and right directions while our proposed descriptor incorporate this information. Our proposed descriptor differentiate multiple directions and invariance to global motion information is addressed at the classification level.

2.4 MOTION DESCRIPTOR

After computing velocity orientation histogram, we then compute statistical features that will capture important motion information. Let n represents the number of frames; we then compute statistics from the time series histogram of velocity orientations. Let $Ht(\omega)$ represents the temporal histogram. We compute following features by using $Ht(\omega)$:
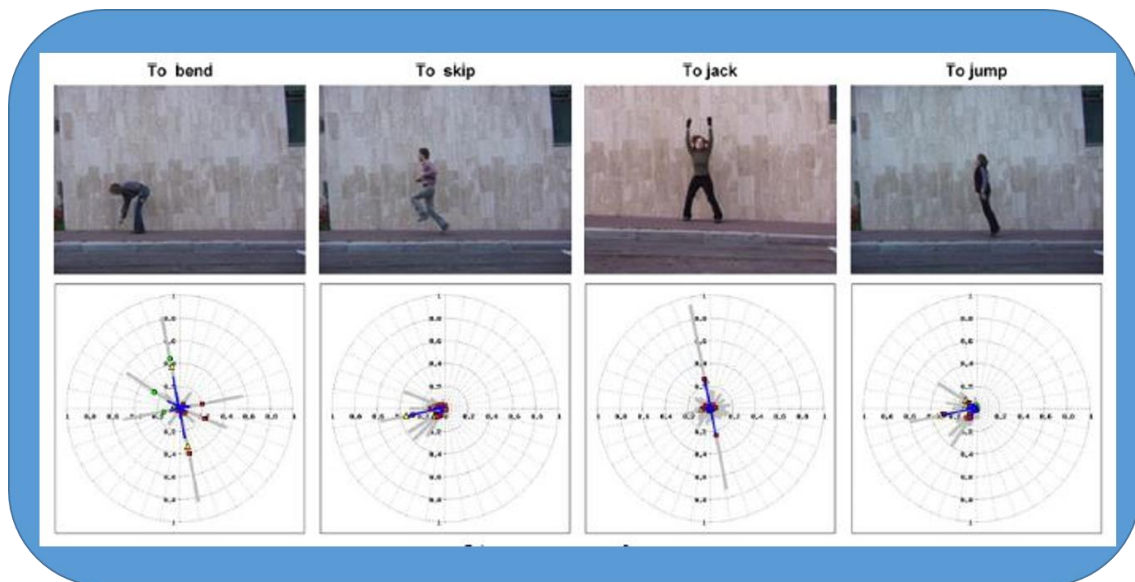

Figure 1: Examples of motion descriptor for different human behaviors

1. We compute the maximum value from the time series of histogram bins $H_t(\omega)$ in the following way

$$\max_{\{0 \le t < n\}}\{H_t(\omega)\}$$

18

2. Mean of $H_t(\omega)$ is computed as

$$\sum_{\{0 \leq t < n\}} \frac{H_t(\omega)}{n}$$

3. Standard deviation of $H_t(\omega)$ is computed as

$$\sigma(\omega) = \sqrt{\sum_{\{0 \leq t < n\}} \frac{H_t{}^2(\omega)}{n} - \mu(\omega)^2}$$

We also divide the time interval into three segments, and compute the corresponding mean as follows. Let n is the total time interval

4. Mean for the first interval $\Psi_a(\omega)$ is formulated as

$$\sum_{\{0 \leq t < \frac{n}{3}\}} \frac{H_t(\omega)}{n/3}$$

5. Mean for the second interval $\Psi_b(\omega)$ is given by

$$\sum_{\{\frac{n}{3} \leq t < \frac{2n}{3}\}} \frac{H_t(\omega)}{n/3}$$

6. Mean for the third interval $\Psi_c(\omega)$ is given by

$$\sum_{\{\frac{2n}{3} \leq t < n\}} \frac{H_t(\omega)}{n/3}$$

Some examples of human behaviors and their corresponding motion descriptor are shown in Figure 1. Before computing motion descriptor, we divide the video into three temporal segments, i.e., and beginning. Middle and end segments [53]. We then compute motion descriptor for each segment of the video and concatenate all the three motion descriptors to represents over all motion behavior. For each motion descriptor, the blue lines represents the maximum values while gray lines represent the mean values. The red square represents mean values for the beginning of the video sequence; yellow triangle represents means values for the middle of the sequence while green disk represents the mean values associated to the end of video sequence [54]. From the Figure 1, it is obvious that our proposed motion descriptor is very discriminative and has the ability to identify distinct human behaviors.

2.5 NON-LINEAR SVM CLASSIFICATION

The SVM classifier applied in many pattern recognition problems. For classification, we use a non-linear support vector machine with a multi-channel kernel that efficiently combines multiple channels. We then define the multi-channel Gaussian kernel by:

$$K(H_i, H_j) = \exp\left(-\sum_{c \in C} \frac{1}{A_c} D_c(H_i, H_j)\right)$$

Where Hi = { hin } and Hj = { hjn } are the histograms for channel c and Dc (Hi, Hj) is the L2 distance defined as

$$D_c(H_i, H_j) = \frac{1}{2} \sum_{n=1}^{V} \frac{(h_{in} - h_{jn})^2}{h_{in} + h_{jn}}$$

Where V is the size of vocabulary. The parameter Ac is the average distances between all training samples for a channel c. For a given training set, we find the best set of channels C based on a greedy approach. We start with empty set of channels and add all possible channels. We then use

greedy approach to evaluate each channel and remove channels until maximum is reached. In the case of multi-class classification, we use an approach of one-against all.

3. EXPERIMENTAL RESULTS

For the performance evaluation, we use dataset in [55]. The video database, as shows in Figure 2, is public available sequences, which contains 93 sequences of 10 human actions. The actions are: bend, jump, jack, jump forward- on-two-legs, jump-in-place-on-two-legs, run, gallop sideways, skip, walk, wave-two-hands, and wave one-hand) performed by 9 different actors. All the video sequences have the resolution of 180x144 pixels and are four seconds long video with average of 50 fps. This data set also include extracted foreground, obtained by background subtraction. We compute optical flow for only the foreground objects excluding the background in order to reduce the computation time.

For training, we use 2/3 of the dataset and the rest for testing. During our experiment, we randomly select six sequences in each action as a training set and the rest for testing.
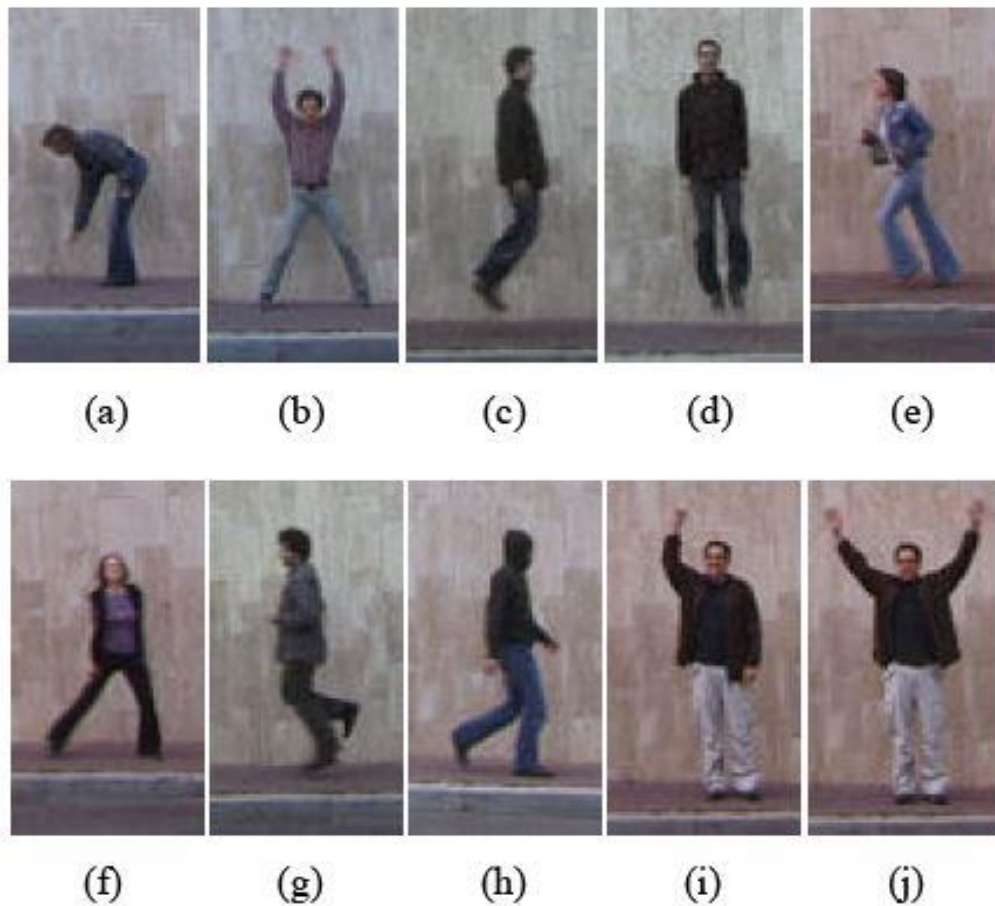

Figure 2. Sample frames from the dataset contain that 10 actions of 9 persons consists of (a) bend, (b) jack, (c) jump, (d) jump, (e) run, f) side, (g) skip, (h) walk, (i) wave1, (j) wave2

Figure 3, shows that confusion matrix for classification, where we train the classifier on one behavior and test it on the other behaviors. The average classification rate is 79.17%. There are some misclassified of bend with pjump because the person stand still before and after bending down which motion vectors is similar to the wrong category. In skipping, 3/4 sequences are

classified as running. It is not unexpected because their motion and posture are very likely to each other. In addition, the misclassification of pjump with jump and wave2 with wave1 cause by similar pose too.

	Bend	Jack	Jump	Pjump	Run	Side	Skip	Walk	Wave1	Wave2
Bend	0.87	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00
Jack	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Jump	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pjump	0.00	0.00	0.33	0.67	0.00	0.00	0.00	0.00	0.00	0.00
Run	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
Side	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
Skip	0.00	0.00	0.00	0.00	0.75	0.00	0.74	0.00	0.00	0.00
Walk	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
Wave1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
Wave2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.67	0.33

We also compare our method with other reference methods and the results are reported in Table 1. The other Classification methods perform leave-one out with nearest neighbor while we used hold out method. In our method, the sequences in training set are not used in the testing process [56]. For Leave-one-out with X samples, the method is train all data except for one sample and test the prediction with the sample in each time X. The average error of X time is computed. So every data used to be a testing once and be a training $X - 1$ times. The variance of resulting evaluation also reduces as a number of training set increases. From the Table 1, it is obvious that our method outperforms other state-of-the-art methods.

METHODS	ERROR RATE
Khan et al [24]	**79.41**
Ullah et al [43]	**95.24**
Kong et al [3]	**85.69**
Proposed	**72.11**

4. CONCLUSION

In this paper, we proposed an approach for recognizing human behaviors using our proposed features and non-linear SVM classifier. We demonstrated the capability of our approach in capturing the the dynamics of different classes by extracting these features. These features adopt the SVM to learn different classes. The main advantage of the proposed method is its simplicity and robustness.

REFERENCES

[1] Wang, Limin, Yuanjun Xiong, Zhe Wang, Yu Qiao, Dahua Lin, Xiaoou Tang, and Luc Van Gool. "Temporal segment networks: Towards good practices for deep action recognition." In European Conference on Computer Vision, pp. 20-36. Springer, Cham, 2016.

[2] Feichtenhofer, Christoph, Axel Pinz, and Richard P. Wildes. "Spatiotemporal multiplier networks for video action recognition." In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7445-7454. IEEE, 2017.

[3] Kong, Yu, Shangqian Gao, Bin Sun, and Yun Fu. "Action Prediction From Videos via Memorizing Hard-to-Predict Samples." In AAAI. 2018.

[4] Ma, Shugao, Leonid Sigal, and Stan Sclaroff. "Learning activity progression in lstms for activity detection and early detection." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1942-1950. 2016.

[5] Hu, Weiming, Dan Xie, Zhouyu Fu, Wenrong Zeng, and Steve Maybank. "Semantic-based surveillance video retrieval." IEEE Transactions on image processing 16, no. 4 (2007): 1168-1181.

[6] Ben-Arie, Jezekiel, Zhiqian Wang, Purvin Pandit, and Shyamsundar Rajaram. "Human activity recognition using multidimensional indexing." IEEE Transactions on Pattern Analysis & Machine Intelligence 8 (2002): 1091-1104.

[7] Saqib, Muhammad, Sultan Daud Khan, and Michael Blumenstein. "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, 2016.

[8] Cutler, Ross, and Larry S. Davis. "Robust real-time periodic motion detection, analysis, and applications." IEEE Transactions on Pattern Analysis and Machine Intelligence 22, no. 8 (2000): 781-796.

[9] Efros, Alexei A., Alexander C. Berg, Greg Mori, and Jitendra Malik. "Recognizing action at a distance." In null, p. 726. IEEE, 2003.

[10] Fathi, Alireza, and Greg Mori. "Action recognition by learning mid-level motion features." In Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, pp. 1-8. IEEE, 2008.

[11] Chaudhry, Rizwan, Avinash Ravichandran, Gregory Hager, and René Vidal. "Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions." In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp. 1932-1939. IEEE, 2009.

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

[13] Khan, Wilayat, Habib Ullah, 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..

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

[15] 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.

[16] 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.

[17] Khan, Sultan Daud, Muhammad Tayyab, Muhammad Khurram Amin, Akram Nour, Anas Basalamah, Saleh Basalamah, and Sohaib Ahmad Khan. "Towards a Crowd Analytic Framework For Crowd Management in Majid-al-Haram." arXiv preprint arXiv:1709.05952 (2017).

[18] 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.

[19] 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.

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

[21] Arif, Muhammad, Sultan Daud, and Saleh Basalamah. "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.

[22] 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.

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

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

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

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

[27] 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.

[28] Khan, Sultan Daud, Giuseppe Vizzari, Stefania Bandini, and Saleh Basalamah. "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.

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

[30] 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).

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

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] 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.

[39] 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.

[40] 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.

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

[42] 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.

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

[44] 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.

[46] Khan, A. (2015). Coordinated Unmanned Aerial Vehicles for Surveillance of Targets (Doctoral dissertation, Queen Mary University of London).

[47] 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.

[48] 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.

[49] Ahmad, K., Conci, N., & De Natale, F. G. (2018). A saliency-based approach to event recognition. Signal Processing: Image Communication, 60, 42-51.

[50] 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.

[51] 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.

[52] 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.

[53] 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.

[54] 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.

[55] 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.

[56] 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.