Intelligent Automation & Soft Computing
DOI: 10.32604/iasc.2022.025013
Article

Tech Science Press

Multiple Events Detection Using Context-Intelligence Features

Yazeed Yasin Ghadi[1], Israr Akhter[2], Suliman A. Alsuhibany[3], Tamara al Shloul[4], Ahmad Jalal[2] and Kibum Kim[5,*]

[1]Department of Computer Science and Software Engineering, Al Ain University, Al Ain, 15551, UAE
[2]Department of Computer Science, Air University, Islamabad, Pakistan
[3]Department of Computer Science, College of Computer, Qassim University, Buraydah, 51452, Saudi Arabia
[4]Department of Humanities and Social Science, Al Ain University, Al Ain, 15551, UAE
[5]Department of Human-Computer Interaction, Hanyang University, Ansan, 15588, Korea
*Corresponding Author: Kibum Kim. Email: kibum@hanyang.ac.kr

Received: 08 November 2021; Accepted: 13 January 2022

Abstract: Event detection systems are mainly used to observe and monitor human behavior via red green blue (RGB) images and videos. Event detection using RGB images is one of the challenging tasks of the current era. Human detection, position and orientation of human body parts in RGB images is a critical phase for numerous systems models. In this research article, the detection of human body parts by extracting context-aware energy features for event recognition is described. For this, silhouette extraction, estimation of human body parts, and context-aware features are extracted. To optimize the context-intelligence vector, we applied an artificial intelligence-based self-organized map (SOM) while a genetic algorithm (GA) is applied for multiple event detection. The experimental results on challenging RGB images and video-based datasets were promising. Three datasets were used. Event recognition and body parts detection accuracy rates for the University of central Florida's (UCF) dataset were 88.88% and 86.75% respectively. 90.0% and 87.37% for event recognition and body parts detection were achieved over the University of Texas (UT) dataset. 87.89% and 85.87% for event recognition and body parts detection were achieved for the sports videos in the wild (SVW) dataset. The proposed system performs better than other current state-of-the-art approaches in terms of body parts and event detection and recognition outcomes.

Keywords: Body parts detection; event detection; context-intelligence features; genetic algorithm; machine learning; self-organized map

1 Introduction

Social interaction and event detection [1] cause a social communication network between millions of posts every minute, and these graphs are growing daily [2]. User-generated information is associated with private or shared interactions, which can be described as multimedia information that can be captured as digital data [3]. Current research studies show that multimedia information is organized based on underlying experiences that facilitate efficient descriptions, synchronization, analysis, indexing, and surfing [4]. Event recognition is an important research area and it is widely used in various applications where high-level image recognition is required, e.g., safety control, smart systems [5–7], data security [8], emergency systems [9], monitoring of interactions between humans and computers [10], intelligent indexing [11] and sports event detection [12]. Regarding the identification of security-related data and events [13], a surveillance footage feature [14] is available in several areas such as smart homes [15], parking spaces, hospitals [16] as well as community sites [17].

In this research study, we describe a new robust computational intelligent system for multiple event detection and classification with the help of a context-intelligent features extraction approach, data optimization and a genetic algorithm. In our approach, we take RGB images and three publicly available video-based datasets, namely, the UCF sports action dataset, the UT-interaction dataset, and the Sports videos in the wild (SVW) dataset. Initially, RGB conversion, noise reduction and binary conversion are performed to minimize the computational cost in time and processing loads. After this, the next step is to extract the human silhouette and detect human body parts with the help of skin detection and Otsu's method. Then, context-intelligent features extraction is performed in which we extract angular geometric features, multi-angle joints features, triangular area points features, the distance between 2 points, and an energy feature. To reduce the computational power burden and to make the system more intelligent, we adopt a machine learning-based data optimization method with the help of a self-organized map (SOM). Finally, for multiple event detection and classification, a genetic algorithm (GA) is adopted. The key contributions of this paper are as follows:

Human silhouette extraction using two different approaches to optimize the human silhouette. Two-dimensional (2D) stick modelling is adopted for human posture information and analyses of human body movement in RGB images and video data. Context-intelligent features extraction in which angular geometric features, multi-angle joint features, triangular area points features, the distance between two points and energy features are extracted. To save time and computational cost, a data optimization technique is adopted while, for multiple event detection, an artificial intelligence-based genetic algorithm is applied.

The majority of the article is arranged as follows: Section 2 explains related work. Section 3 describes the layout of the system and displays our conceptual system architecture which involves a pre-classification process that explains individual object segmentation, initialization of body parts, the identification of eight body parts, and the extraction of distance and energy features. Section 4 discusses our hypotheses and explains the quality of our system in three separate tables. Section 5 offers a conclusion and a note on future work.

2 Related Work

Several research studies have documented their efforts to detect vital and informative movements of human body points. In [18], Einfalt et al. established a two-step system for extracting sequential 2D pose configurations from videos for event identification in the movement of players. Using localization activity classification, they created a convolution layers segment network to specifically recognize such events. Their procedure outlines skin tone detection with a green (Y), blue (Cb), red (Cr) YCbCr color model, heuristic [19] thresholds and skin tone improvement [20]. Jalal et al. [21] worked on human activity detection using video depth without adding any movement. They developed a system of random forest iteration with specific temporal characteristics to demonstrate different actions [22]. In [23] they presented an alternative method that identified human body part silhouettes using Hidden Markov Models. To track human pose, Lee et al. [24] introduced an innovative hierarchical system that uses edge-based functionality in the rough stage. Aggarwal et al. [25] designed Human movement analysis using 2D and three-dimensional (3D) shape analysis. Wang et al. [26] built a framework for estimating human movement and for recognizing behaviors. In [27] authors proposed to integrate neural network models with the conceptual hierarchy with human bodies. As a result, they described the strategy as a structure for combining neural information. This framework is easy to put together the results of several inference procedures across a period of straight reasoning (directly anticipating each component of the system) lower part prediction (using visual data to estimate a human body), (building information from disparate components), and assumption from the bottom up. The lowest part and upper assumptions, including both, show up the compositional connections in human bodies. In authors proposed a heuristic approach for the detection of human-object interaction via human-based video and images data. In [28] author used Graph Parsing Neural Network (GPNN), which includes i) the graph structure and adjacency matrix, and ii) the labeled node. They used the Pixel Value Co-occurrence Model (GLCM) as well as the Local Binary Pattern (LBP) to build a novel mixed features descriptor technique for intensive detections LBP. For activity recognition, researchers combined obtained features with pattern recognition supervised classification techniques. In [29] authors described a detailed analysis that covers a wide range of topics, including algorithm taxonomy to unresolved problems. They initially look into deep salient object detection (SOD) methods from a variety of angles, covering network design, supervisory level, learning methodology, and instrument identification. The current SOD statistics and assessment metrics are then summarized. Furthermore, they compare a wide number of sample SOD systems and give in-depth assessments of the outcomes. Additionally, they create a unique SOD dataset containing extensive characteristic descriptions spanning multiple object classification kinds, difficult aspects, and scene classifications to examine the behavior of SOD algorithms with varied characteristic configurations, which have not been properly examined previously. In [30] the researcher proposed a new approach for human gaze communication in public videos that are studied at both the atomic and the event levels, which are important for understanding human social interactions. To address this unique and difficult challenge, we provide vacation, a multimedia content dataset that includes comprehensive descriptions of objects including human faces, human engagement, and communication structures and labeling at both the atomic and the event levels.

In this research article, we propose a vital method for event detection in which a salient area detection process is applied to detect visibly significant regions and skin tone detection is implemented for background segmentation. For the body parts model, eight main points of the body are identified, and a 2D stick model is constructed and implemented. After human detection, the next step is feature extraction of the detected human silhouette. Two types of feature extraction methods are performed on the UCF sports action dataset, the first is the distance between body parts feature and the second is the context-aware energy feature.

3 Material and Methods

In the designed system methodology, we elaborate on our human event recognition (HER) in the following phases (1) pre-processing, (2) human detection and silhouette extraction, (3) skeleton and 2D stick model (4) feature extraction, and (5) classification. The design architecture of the proposed system is shown in Fig. 1.

Primarily, for silhouette identification and segmentation processes, we require multiple steps such as the detection of height and width of associated parts, skin identification, noise reduction, and ambient silhouette separation. For the detection of skin, a tone and connected components method is applied; eight human body parts are detected via skin pixel and the resizing of images, then 2D stick modeling of human skeletons is performed upon the extracted human body points. After that, we extract the energy and distance features from the RGB images. Next, we find the event class of the data by applying SOM as a data optimizer and the GA algorithm for event classification.

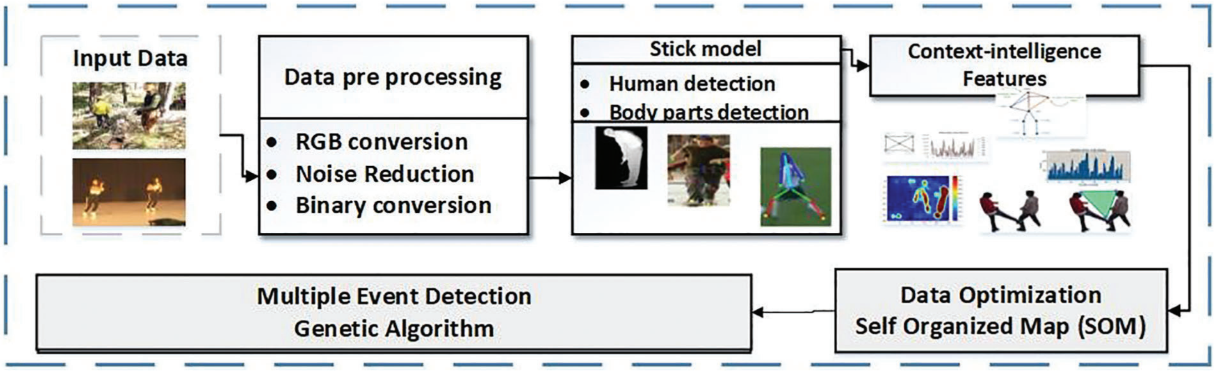


Figure 1: System architecture of the proposed system

3.1 Pre-Processing

During the pre-processing of RGB images, resizing of the images, noise removing via Gaussian filter [31,32] and RGB to binary (0, 1) image conversion [33,34] are applied.

3.2 Silhouette Extraction

In this section, we describe two methods for silhouette extraction. Firstly, we have a human skin detection method to find skin pixels from the images [35,36]. A Gaussian filter is applied to remove noise in images where the skin tone is detected and for the segmentation of [37,38] human silhouettes we use heuristic [39] thresholding [40]. In step 1, we define a heuristic threshold [41] value δ (see Eq. (1)) with an existing Otsu's image thresholding method [42,43] which is represented as;

$$\delta = \begin{cases} R\left(\dfrac{\text{ThO} + \text{Thmax}}{4}\right), & \textit{if Thmax} \le 10, \\ R\left(\dfrac{\text{ThO} + \text{Thmax}}{2}\right), & \textit{if Thmax} > 10, \end{cases} \tag{1}$$

where R is round, ThO is the threshold defined by Otsu's method and $Thmax$ is the maximum point of color frequency for defined histogram values. This process is implemented for each greyscale region of an image, formulated as Eq. (2);

$$I(RGB) = \begin{cases} (0,\ 0,\ 0), & \textit{if } g(x,\ y) = 0, \\ (Ir(x,\ y),\ Ig(x,\ y),\ Ib), & \textit{if } g(x,\ y) = 1, \end{cases} \tag{2}$$

Then, we extract the skin pixel region from the human silhouette via a skin detection technique in which YCbCr is used to identify the skin tone regions [44,45]. In the second step; a propagation-based method for saliency detection [46,47] is applied. These skin detection and propagation-based methods for saliency detection are merged for further processing [48,49]. Fig. 2 shows silhouette representation on saliency detection and propagation methods.

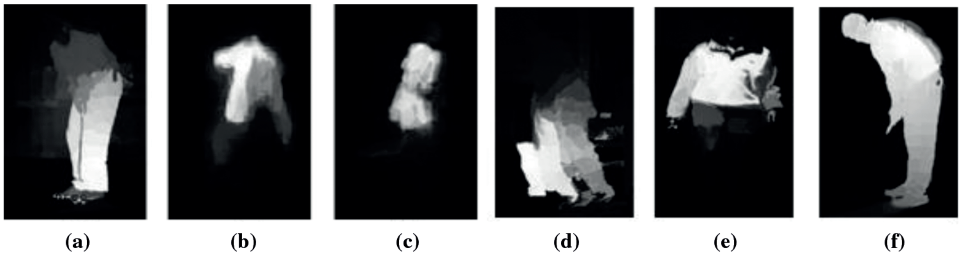


Figure 2: Examples of silhouette representation. (a) Golf side view, (b) Kick backside view, (c) Running and kicking, (d) Football kick, (e) Walk and (f) Golf side view

3.3 Silhouette Extraction

Once the silhouette is obtained, we initialize the human silhouette to implement the algorithm for the body parts. In body parts extraction [50–52] we detect the general body parts using skin algorithms, human body shape, and angle techniques.

3.4 Key Body-Parts Model

In this section, we provide a detailed overview of our key body parts [53] model. Initially, we select 8 points on the human body, namely, head point, torso, right/left hands, right/left knees, and right/left feet. We detect these points by applying skin detection and image resizing methods. For head point detection, we find skin pixels in the image using skin tone detection. In this technique, the binary image is used and the search is started from the top to the head position. The following formulation Eq. (3) is used for head tracking;

$$K_H^I \leftarrow K_H^{I-1} + \Delta K_H^{I-1}, \tag{3}$$

where KH^{I} is the location of the head point at any particular frame *I*. This is attained to find a correlation in the arrangements of frames. For torso detection, we take the mid-point of all the skin pixels. For right-hand detection, we use the hand width and height model to find the side of the hand. After hand detection, our model detects knee and footpoints. After this, we applied a 2D stick model from our detected human body parts.

3.5 2D Stick-model

After body initialization, we describe the stick model which consists of 7 sticks that are connected between the body points to represent [54–56] the human skeleton. Head, neck, shoulders and hand points are considered as upper 2D stick maps, while feet, knees, and torso points are considered as lower 2D stick maps [57]. The head point is connected with the hands and torso point. Fig. 3 represents the 2D stick model.

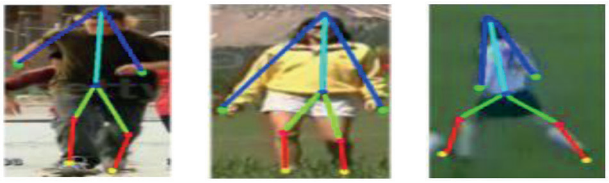


Figure 3: 2D Stick model examples on three different actions

3.6 Feature Representation

3.6.1 Angular Geometric Features

From this type of feature, we consider the areas of the triangles as angular geometric features. We have triangle one (head, right hand, and mid-point), triangle two (head, left hand, and mid-point), triangle three (mid, right knee, and right foot), triangle four (mid, left knee, and left foot). Eq. (4) shows the mathematical representation of the angular geometric features

$$A_{gf} = 1/2\{Ha_1(L_{hp2-}M_{p2}) + Lhp_1(M_{p2-}\ Ha_2) + Mp_1(Ha_2 - L_{hp2})\}, \tag{4}$$

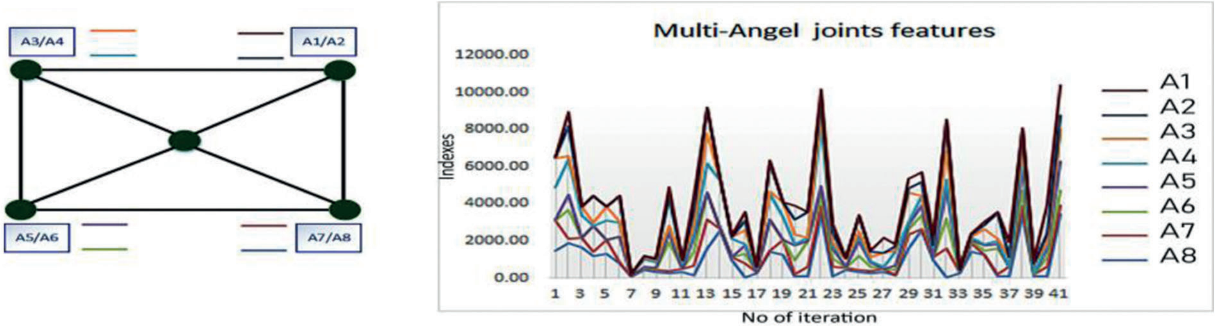
where A_gf denotes the triangular area, 〖*Ha*〗 _1 and 〖*Ha*〗 _2 show the head points, L_(hp2) and L_hp1 show the left-hand points and 〖*Mp*〗 _1, 〖*Mp*〗 _2 denote midpoints of the human body.

3.6.2 Multi-Angle Joints Features

For multi-angle joint features, we apply this procedure over detected human body parts. A 5 × 5 pixel-based window is defined by considering the center pixel of each detected human body part. After that, we derive eight angles by developing four triangles to find the angle information. Eq. (5) shows the mathematical information:

$$\begin{aligned} &A1 = \cos(i,\, j) \rightarrow l,\; A2 = \cos(i,\, j) \rightarrow l,\; A3 = \cos(i,\, j) \rightarrow l,\; A4 = \cos(i,\, j) \rightarrow l, \\ &A5 = \cos(i,\, j) \rightarrow l,\; A6 = \cos(i,\, j) \rightarrow l,\; A7 = \cos(i,\, j) \rightarrow l,\; A8 = \cos(i,\, j) \rightarrow l, \end{aligned} \tag{5}$$

where $A1$, $A2$, $A3$, $A4$, $A5$, $A6$, $A7$ *and* $A8$ show the edges of the four triangles and $\cos(i, j)$ shows the angle value of giving (i, j) pixels, and $\rightarrow l$ for the sides of the windows. Fig. 4 shows the results and conceptual design for the multi-angle joint features.



(a) **(b)**

Figure 4: (a) Conceptual design and (b) Results of multi-angle joint features

3.6.3 Triangular Area Points Features

In triangular area points features, we considered a triangular shape over the detected human body parts. For this we connected the head points of two people in the interaction and connected the point of contact of both people as point three (See Fig. 5). After this, we find the area with the help of Eq. (6).

$$T_{ap} = 1/2\{Ha_1(H_{b2-}C_{p2}) + H_{b1}(C_{p2-}\ Ha_2) + Cp_1(Ha_2 - H_{b2})\}, \tag{6}$$

where T_{ap} denotes the triangular point area, Ha_1 *and* Ha_2 shows the head points of person one, H_{b2} and H_{b1} shows the head point of person two and Cp_1, Cp_2 denotes the first connection point of both persons. Fig. 5 shows the complete overview of the triangular area point features.

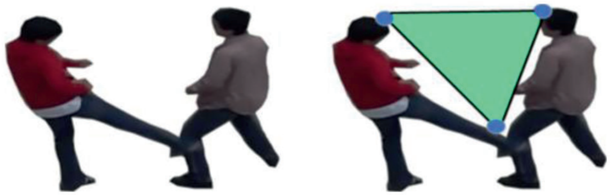


Figure 5: Complete overview of triangular area point features

3.6.4 Distance Measuring Between Two Points

In this section, we find the distance between each set of two points, namely, head to torso-point, head to hands, torso points to knee and knee to footpoints. The distance between two edge points $b1$ and $b2$ having x, the y coordinates are given as Eq. (7);

$$Distance(P1\ ,\ \ P2) = \sqrt{(P1x - P2x)^2 + (P1y - P2y)^2}, \tag{7}$$

where $Distance(P1\ ,P2)$ is the Euclidean distance.

3.6.5 Energy Feature Representation

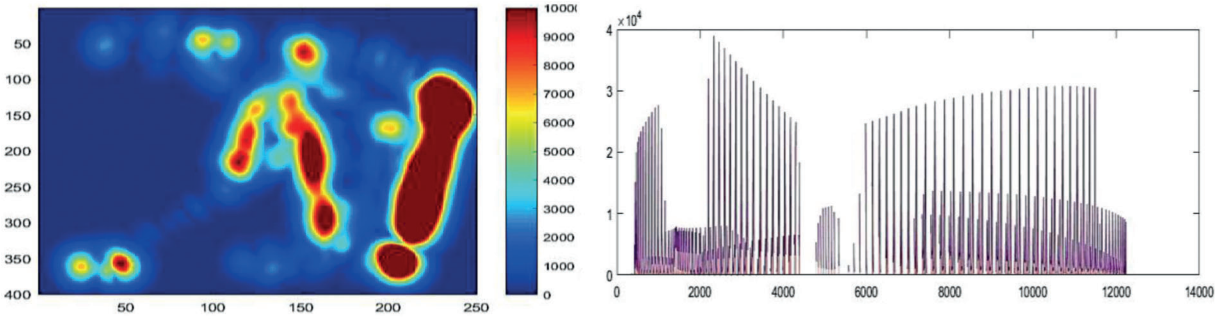
In the energy feature section, we extract context-aware energy features over RGB images by applying an energy map to the entire image. Initially, we examine an energy index-based matrix with the range of 0–10000 indexes which is based on each silhouette. We find a certain threshold index value and find the RGB value of a particular index from the energy map matrix and store them in a vector. Eq. (8) shows the energy vector and Fig. 6 shows the result for energy features extraction.

$$Eng = \sum_{0}^{w} InR(w), \tag{8}$$

where Eng is denoted as an energy vector and w shows the index values and InR represents the RGB values of a particular index pixel. After getting the energy vector, we concatenate it with the distance vector for further classification. Eq. (9) represents the context-intelligent feature vector as;

$$FV = Distance(P1\ ,\ P2)\ \&\ Eng\ \&\ T_{ap}\ \&\ A1\_A8\ \&\ A_{gf}, \tag{9}$$

where FV is context-intelligent features vector, $Distance(P1\ ,P2)$ is the distance, Eng is energy feature vector, T_{ap} is the triangular area points features, $A1\_A8$ is the multi-angle joint features and A_{gf} is the Angular geometric feature.



(a) (b)

Figure. 6: (a) Energy features results and (b) Energy vector 1D representation

3.7 Events Classification

For event classification, we used two machine learning classifiers: SOM (Self-organized map) [58] as a pre-classifier and (GA) genetic algorithm as a classifier.

3.7.1 Self-organized map (SOM)

A self-organizing map (SOM) [59] is trained by an unsupervised learning method [60] in which training samples are discretized representations of the input that is called a map [61]. Self-organizing maps apply competitive learning methods as back propagation with gradient descent, [62] during pre-classification [63] the SOM algorithm shapes a map between the high-dimensional data space of a typical two-dimensional data structure [64]. The model vectors are situated in the data space which acts as an ordered set of different types of data items [65]. The map is used as an ordered groundwork for illustrating different aspects of the dataset [66]. After that, we applied the artificial neural networks (ANN) to get better results [67].

3.7.2 Genetic Algorithm (GA)

A genetic algorithm (GA) [68] is adopted as a prediction model for event detection, [69] identification [70], and recognition [71]. Three general steps [72–74] are adopted in the genetic algorithm [75] for each interval to produce the upcoming generation with the help of the existing population [76]. Initially, the selection of the individual's forms [77] given data which are called parent nodes [78–80]. Parents are essential for the next or upcoming generation [81]. For the next solution [82], the base chromosome follows a cross-over step over children's, as represented in Eq. (10).

$$Cfit(Cr) = \frac{fj}{\sum_{j=0}^{n} fj}, \tag{10}$$

where $Cfit(Cr)$ denotes the fitness function over j^{th} iteration and fj are the chromosome fitness values [83–86]. Finally, mutation procedures are performed to find the optimal solution from the given data.

4 Performance Evaluation

This section describes the detailed overview of three datasets [87–89] used for UCF Sports dataset, UT-interaction, and Sports Videos in the Wild (SVW) dataset. Various experimental results evaluate for the proposed system with other state-of-the-art systems.

4.1 Datasets Description

Three benchmark datasets have been used, i.e., UCF Sports dataset, UT-interaction, and Sports Videos in the Wild (SVW). Tab. 1 shows the detailed description of datasets

Table 1: Datasets description in detail that have been used in experimental setup

Datasets name	Datasets input	Description
UCF Sports dataset	RGB videos	The UCF sports dataset [90] contains several acts including jumping, swinging golf, punching, walking, riding horse, biking, skateboarding, swinging bench and swing side along with human annotations.
UT-interaction	RGB videos	The UT-interaction dataset consists of six classes which are based on videos of one-minute duration. Handshake, human pointing, embrace, drive, strike and kick are the six classes. The dataset frame rate is 30fps, and the frame size is 720 × 480 pixels.
Sports videos in the wild	RGB videos	Sports Videos in the Wild (SVW) [91] an application is a TechSmith organization-based product. There are nineteen event-based classes.

4.2 Experimental Settings and Results

4.2.1 Experimental Results on Datasets

Experiment I: Human Body Parts Detection

To test the efficacy of our proposed system, we first measured the Euclidean distance from the ground truth of each identified body part and our proposed system. To compute the Euclidean distance, the formula and Eq. (11) is:

$$\boldsymbol{Distance_P} = \sqrt{\sum_{P=1}^{P}\left(\frac{Xp}{Sp} - \frac{Y_p}{S_p}\right)^2}, \tag{11}$$

where xp is the ground truth for human body parts detection, the point yp is identified and the distance which is found by error calculator, $\boldsymbol{Distance_p}$ is the distance Euclidean. To analyze the ground truth of our identified point technique, we used 20 pixels as a margin of an error value. Tab. 1 presents the performance accuracy of human body key point recognition and shows experimental results for the UCF Sports action dataset, the UT interaction dataset and the sports videos in the wild (SVW) dataset. Our system achieved accuracy for recognition of human body parts of 86.67% for the UCF Sports action, 87.37% for the UT interaction dataset, and 85.87% for the sports videos in the wild (SVW) dataset as shown in Tab. 2.

Table 2: Results for the detection of human body parts over the UCF, UT and SVW dataset

Body parts	Distance	UCF	Distance	UT	Distance	SVW
Mean detection accuracy		86.75%		87.37%		85.87%

Experiment II: Multiple Event Detection

For multiple event detection, after the data optimization, artificial intelligence based on a computational intelligent genetic algorithm is applied. Tab. 3 shows the confusion matrix with 90.00% accuracy for event recognition over the UT-interaction dataset. Tab. 4 shows the confusion matrix with 88.88% accuracy for event recognition over the UCF sports action dataset. Tab. 5 shows the confusion matrix with 87.89% accuracy for event recognition over the SWV dataset.

Table 3: Confusion matrix of GA for event recognition over the UT_interaction dataset

	HS	HG	KI	PO	PU	PS	Mean accuracy
HS	9	8	9	10	10	8	90.00%

Note: HS = hand_shaking, HG = hugging, KI = kicking, PO = pointing, PU = punching, PS = pusing.

Table 4: Confusion matrix of GA for event recognition over the SVW dataset

AR	BB	BT	BM	BW	CD	FB	GL	HJ	HK	HR	JV	LJ	PT	RW	ST	SK	TN	VL
9	8	8	9	9	9	9	8	8	9	10	10	9	9	10	8	8	9	8
Mean event detection accuracy = 87.89%																		

Note: AR = archery, BB = baseball, BT = basketball, BM = bmx, BW = bowling, CD = cheerleading, FB = football, GL = golf, HJ = highjump, HK = hockey, HR = hurdling, JV = javelin, LJ = longjump, PT = polevault, RW = rowing, ST = shotput, SK = skating, TN = tennis, VL = volleyball.

Table 5: Confusion matrix of GA for event recognition over the UCF dataset

FK	GFS	GS	HR	KK	RS	SB	W1	W	Mean accuracy
8	9	10	9	8	9	9	9	9	**88.88%**

Note: FK = Front kick, GFS = Golf Front side, GS = Golf side, HR = Horse riding, KK = Karate kick, RS = Running Side, SB = Skate boarding, W1 = Walk-1, W = walk.

Experiment III: Comparison with Other Classification Algorithms

In this phase, we evaluate the precision, recall, and f-1 measure over the UCF sports action dataset, the SVW dataset, and the UT-interaction dataset. For the classification [92] of multiple event comparison, we used the machine learning and artificial intelligence-based Genetic algorithm (GA), Artificial Neural Network (ANN) and Adaboost. Fig. 7 shows the comparison of machine learning classifiers for precision, recall and F-1 measure over the UT-interaction dataset.

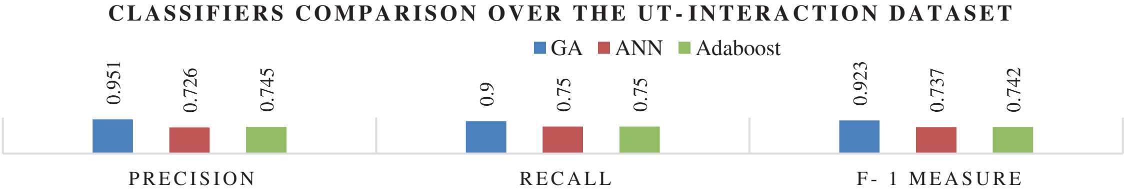


Figure. 7: Machine learning classifiers comparisons for precision, recall and F-1 measure over the UT-interaction dataset

4.2.2 Comparison with Other Systems

Comparisons of our proposed system with state-of-the-art methods [93], as shown in Tab. 6, indicate that our efficiency [94] on the datasets is much better [95] the existing approaches listed in Tab. 6. The Markov random field model is used by Park et al. [96]. Which combine pixels into linked blobs and to record inter-blob relations. Traditional neural networks are used by Li et al. [97] to estimate human body pose. H. W. Chen et al. [98] used morphological segmentation of the surface color and systematic thresholding. Rodriguez et al. proposed a novel method for estimating future body motion. They used realistic explanations and targeted failure processes to motivate a reproductive system to predict specific future human motion. Tab. 5 shows the detailed multiple event detection and classification comparison with state-of-the-art methods and techniques.

Table 6: Event classification comparison between the proposed method and state-of-the-art methods

Methods	UCF	Methods	UT	Methods	SVW
Park et al. [96]	64.33	Rodriguez et al. [99]	71.80	Sun et al. [102]	74.20
Li et al. [97]	78.22	Xing et al. [100]	85.67	Rachmadi et al. [103]	82.30
Chen et al. [98]	79.0	Chattopadhyay et al. [101]	89.25	Zhu et al. [104]	83.10
Proposed method	**88.88%**		**90.0%**		**87.89%**

5 Research Limitation

In this paper, the challenging datasets are used. In which, we had small discrepancies in findings due to complicated perspective information and complexity of human data. Occlusion and merging problems in a specific region while working with these aspects of data and situations, we ran across several issues. We will research this challenge in the future and adopt a deep learning technique, and we will create a new strategy to get excellent outcomes.

6 Conclusion

In this paper, we proposed a method for detecting complex human activity-based events with novel context-intelligence features and 2D stick models. To identify body parts and event detection in RGB images, we introduced an enhanced multi-function extraction design. To optimize the context-intelligence vector we applied an artificial intelligence-based self-organized map (SOM). A genetic algorithm (GA) is applied for multiple event detection. The experimental results were obtained through challenging RGB images and three videos-based datasets, namely, the UCF sports action dataset, the UT-interaction dataset, and the sports videos in the wild dataset. The proposed system performs better than current state-of-the-art approaches in terms of body parts and event recognition. In the future, we will introduce some additional features such as intensity vector and color vector to increase the efficiency of our human event recognition (HER) system.

Funding Statement: This research was supported by grants from the Korea Medical Device Development Fund funded by the Korean government (the Ministry of Science and ICT, the Industry and Energy, Health & Welfare, and Food and Drug Safety) (NTIS 9991006786, KMDF_PR_20200901_0113). Also, this research is supported by the Ministry of Culture, Sports and Tourism and Korea Creative Content Agency (Project Number: R2021040093).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] M. Gochoo, S. R. Amna G. Yazeed, A. Jalal, S. Kamal et al., “A Systematic Deep Learning Based Overhead Tracking and Counting System Using RGB-D Remote Cameras,” *Applied Sciences*, vol. 11, pp. 1–21, 2021.

[2] A. Jalal, S. Lee, J. T. Kim and T. S. Kim, “Human activity recognition via the features of labeled depth body parts,” in *Proc of Int. Conf. on Smart Homes and Health Telematics*, Berlin, Heidelberg, pp. 246–249, 2012.

[3] A. Jalal, M. Batool and S. B. Ud Din Tahir, “Markerless sensors for physical health monitoring system using ecg and gmm feature extraction.” in *Proc of. Conf. on Applied Sciences and Technologies (IBCAST)*, Islamabad, Pakistan, pp. 340–345, 2021.

[4] S. B. ud din Tahir, A. Jalal and K. Kim, “Wearable inertial sensors for daily activity analysis based on adam optimization and the maximum entropy markov model,” *Entropy*, vol. 22, no. 5, pp. 1–19, 2020.

[5] S. A. Rizwan, A. Jalal, M. Gochoo and K. Kim, “Robust active shape model via hierarchical feature extraction with sfs-optimized convolution neural network for invariant human age classification,” *Electronics*, vol. 10, no. 4, pp. 465, 2021.

[6] A. Jalal, Y. Kim and D. Kim, “Ridge body parts features for human pose estimation and recognition from RGB-D video data,” in *Proc of Fifth Int. Conf. on Computing, Communications and Networking Technologies (ICCCNT)*, Hefei, China, pp. 1–6, 2014.

[7] M. A. Ur Rehman, H. Raza and I. Akhter, “Security enhancement of hill cipher by using non-square matrix approach,” in *Proc. Conf. on Knowledge and Innovation in Engineering, Science and Technology*, Berlin, Germany, pp. 1–7, 2018.

[8] A. Jalal, S. Kamal and D. Kim, “Depth map-based human activity tracking and recognition using body joints features and self-organized map,” in *Proc of Fifth Int. Conf. on Computing, Communications and Networking Technologies (ICCCNT)*, Hefei, China, pp. 1–6, 2014.

[9] S. B. Ud Din Tahir, A. Jalal and M. Batool, “Wearable sensors for activity analysis using smo-based random forest over smart home and sports datasets,” in *Proc of 3rd Int. Conf. on Advancements in Computational Sciences (ICACS)*, Lahore, Pakistan, pp. 1–6, 2020.

[10] A. Jalal and Y. Kim, “Dense depth maps-based human pose tracking and recognition in dynamic scenes using ridge data,” in *Proc of 11th IEEE Int. Conf. on Advanced Video and Signal Based Surveillance (AVSS)*, Seoul, South Korea, pp. 119–124, 2014.

[11] A. Jalal and S. Kamal, “Real-time life logging via a depth silhouette-based human activity recognition system for smart home services,” in *Proc of 11th IEEE Int. Conf. on Advanced Video and Signal Based Surveillance (AVSS)*, Seoul, South Korea, pp. 74–80, 2014.

[12] A. Jalal, I. Akhtar and K. Kim, “Human posture estimation and sustainable events classification via pseudo-2d stick model and k-ary tree hashing,” *Sustainability*, vol. 12, no. 23, pp. 9814, 2020.

[13] S. Kamal and A. Jalal, “A hybrid feature extraction approach for human detection, tracking and activity recognition using depth sensors,” *Arabian Journal for Science and Engineering*, vol. 41, no. 3, pp. 1043–1051, 2016.

[14] A. Jalal, Y. -H. Kim, Y. -J. Kim, S. Kamal and D. Kim, “Robust human activity recognition from depth video using spatiotemporal multi-fused features,” *Pattern Recognition*, vol. 61, pp. 295–308, 2017.

[15] A. Jalal, S. Kamal and D. -S. Kim, “Detecting complex 3D human motions with body model low-rank representation for real-time smart activity monitoring system,” *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 12, no. 3, pp. 1189–1204, 2018.

[16] M. Mahmood, A. Jalal and H. A. Evans, “Facial expression recognition in image sequences using 1D transform and gabor wavelet transform,” in *Proc of Int. Conf. on Applied and Engineering Mathematics (ICAEM)*, Islamabad, Pakistan, pp. 1–6, 2018.

[17] N. Khalid, M. Gochoo, A. Jalal and K. Kim, “Modeling two-person segmentation and locomotion for stereoscopic action identification: A sustainable video surveillance system,” *Sustainability*, vol. 13, no. 2, pp. 970, 2021.

[18] M. Einfalt, C. Dampeyrou, D. Zecha and R. Lienhart, “Frame-level event detection in athletics videos with pose-based convolutional sequence networks,” in *Proc of the 2nd Int. Workshop on Multimedia Content Analysis in Sports*, Nice, France, pp. 42–50, 2019.

[19] A. Jalal, S. Kamal and D. Kim, “A depth video-based human detection and activity recognition using multi-features and embedded hidden markov models for health care monitoring systems,” *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. 4, pp. 54–62, 2017.

[20] F. Farooq, J. Ahmed and L. Zheng, “Facial expression recognition using hybrid features and self-organizing maps,” in *Proc of IEEE Int. Conf. on Multimedia and Expo (ICME)*, Hong Kong, pp. 409–414, 2017.

[21] A. Jalal, J. T. Kim and T. -S. Kim, “Development of a life logging system via depth imaging-based human activity recognition for smart homes,” in *Proc of the Int. Symp. on Sustainable Healthy Buildings*, Seoul, Korea, vol. 19, 2012.

[22] A. Jalal, N. Sarif, J. T. Kim and T. -S. Kim, “Human activity recognition via recognized body parts of human depth silhouettes for residents monitoring services at smart home,” *Indoor and Built Environment*, vol. 22, no. 1, pp. 271–279, 2013.

[23] A. Jalal, J. T. Kim and T. -S. Kim, “Human activity recognition using the labeled depth body parts information of depth silhouettes,” in *Proc of the 6th Int. Symp. on Sustainable Healthy Buildings*, Seoul, Korea, vol. 27, 2012.

[24] M. W. Lee and R. Nevatia, “Body part detection for human pose estimation and tracking,” in *Proc of IEEE Workshop on Motion and Video Computing (WMVC’07)*, Austin, TX, USA, pp. 23–30, 2007.

[25] J. K. Aggarwal and Q. Cai, “Human motion analysis: A review,” *Computer Vision and Image Understanding*, vol. 10, pp. 428–440, 1999.

[26] L. Wang, W. Hu and T. Tan, “Recent developments in human motion analysis,” *Pattern Recognition*, vol. 36, no. 3, pp. 585–601, 2003.

[27] W. Wang, Z. Zhang, S. Qi, J. Shen, Y. Pang *et al.,* “Learning compositional neural information fusion for human parsing,” in *Proc. of the IEEE/CVF Int. Conf. on Computer Vision*, Seoul, South Korea, pp. 5703–5713, 2019.

[28] B. H. Lohithashva, V. N. M. Aradhya and D. S. Guru, “Violent video event detection based on integrated LBP and GLCM texture features,” *Revue D’Intelligence Artificielle*, vol. 34, no. 2, pp. 179–187, 2020.

[29] W. Wang, Q. Lai, H. Fu, J. Shen, H. Ling *et al.,* “Salient object detection in the deep learning era: An in-depth survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, pp. 1–20, 2021.

[30] L. Fan, W. Wang, S. Huang, X. Tang and S. -C. Zhu, “Understanding human gaze communication by spatio-temporal graph reasoning,” in *Proc of the IEEE/CVF Int. Conf. on Computer Vision*, Seoul, South Korea, pp. 5724–5733, 2019.

[31] A. Jalal and S. Kamal, “Improved behavior monitoring and classification using cues parameters extraction from camera array images,” *International Journal of Interactive Multimedia & Artificial Intelligence*, vol. 5, no. 5, pp. 1–22, 2019.

[32] A. Jalal, M. A. K. Quaid and A. S. Hasan, “Wearable sensor-based human behavior understanding and recognition in daily life for smart environments,” in *Proc of Int. Conf. on Frontiers of Information Technology (FIT)*, Islamabad, Pakistan, pp. 105–110, 2018.

[33] M. Waheed, A. Jalal, M. Alarfaj, Y. Ghadi, T. Shloul *et al.,* “An LSTM-based approach for understanding human interactions using hybrid feature descriptors over depth sensors,” *IEEE Access*, vol. 10, pp.13, 2021.

[34] Y. Ghadi, I. Akhter, M. Alarfaj, A. Jalal and K. Kim, “Syntactic model-based human body 3D reconstruction and event classification via association based features mining and deep learning,” *PeerJ Computer Science*, vol. 7, pp. e764, 2021.

[35] E. Buza, A. Akagic and S. Omanovic, “Skin detection based on image color segmentation with histogram and k-means clustering,” in *Proc of 10th Int. Conf. on Electrical and Electronics Engineering (ELECO)*, Bursa, Turkey, pp. 1181–1186, 2017.

[36] M. Mahmood, A. Jalal and M. A. Sidduqi, “Robust spatio-temporal features for human interaction recognition via artificial neural network,” in *Proc of Int. Conf. on Frontiers of Information Technology (FIT)*, Islamabad, Pakistan, pp. 218–223, 2018.

[37] I. Akhter, A. Jalal and K. Kim, “Pose estimation and detection for event recognition using sense-aware features and adaboost classifier,” in *Proc of. Conf. on Applied Sciences and Technologies (IBCAST)*, Islamabad, Pakistan, pp. 500–505, 2021.

[38] A. Jalal and M. Mahmood, “Students’ behavior mining in e-learning environment using cognitive processes with information technologies,” *Educational and. Information Technologies*, vol. 24, pp. 2797–2821, 2019.

[39] K. Kim, A. Jalal and M. Mahmood, “Vision-based human activity recognition system using depth silhouettes: A smart home system for monitoring the residents,” *Journal of Electrical Engineering & Technology*, vol. 14, no. 6, pp. 2567–2573, 2019.

[40] P. T. Krishnan, P. Balasubramanian and C. Krishnan, “Segmentation of brain regions by integrating meta heuristic multilevel threshold with markov random field,” *Current Medical Imaging*, vol. 12, pp. 4–12, 2016.

[41] J. K. Leader, B. Zheng, R. Rogers, F. Sciurba, A. Perez *et al.,* “Automated lung segmentation in X-ray computed tomography: Development and evaluation of a heuristic threshold-based scheme,” *Academic Radiology*, vol. 10, pp. 1224–1236, 2003.

[42] N. Amir, A. Jalal and K. Kim, “Automatic human posture estimation for sport activity recognition with robust body parts detection and entropy markov model,” *Multimedia Tools and Applications*, vol. 80, pp. 21465–21498, 2021.

[43] J. Zhang and J. Hu, “Image segmentation based on 2D otsu method with histogram analysis,” in *Proc of Int. Conf. on Computer Science and Software Engineering*, Wuhan, China, vol. 6, pp. 105–108, 2008.

[44] A. Jalal, S. Kamal and D. Kim, “A depth video sensor-based life-logging human activity recognition system for elderly care in smart indoor environments,” *Sensors*, vol. 14, pp. 11735–11759, 2014.

[45] M. Batool, A. Jalal and K. Kim, “Sensors technologies for human activity analysis based on SVM optimized by PSO algorithm,” in *Proc of Int. Conf. on Applied and Engineering Mathematics (ICAEM)*, Islamabad, Pakistan, pp. 145–150, 2019.

[46] A. Farooq, A. Jalal and S. Kamal, "Dense RGB-D map-based human tracking and activity recognition using skin joints features and self-organizing map," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 9, no. 5, pp. 1856–1869, 2015.

[47] M. Pervaiz, Y. Y. Ghadi, M. Gochoo, A. Jalal, S. Kamal *et al.,* "A smart surveillance system for people counting and tracking using particle flow and modified som," *Sustainability*, vol. 13, no. 10, pp. 5367, 2021.

[48] S. Hafeez, A. Jalal and S. Kamal, "Multi-fusion sensors for action recognition based on discriminative motion cues and random forest," in *Int. Conf. on Communication Technologies (ComTech)*, Rawalpindi, Pakistan, pp. 91–96, 2021.

[49] M. Javeed, M. Gochoo, A. Jalal and K. Kim, "HF-SPHR: Hybrid features for sustainable physical healthcare pattern recognition using deep belief networks," *Sustainability*, vol. 13, pp.1699, 2021.

[50] I. Akhter, "Automated posture analysis of gait event detection aia a hierarchical optimization algorithm and pseudo 2D Stick-model," *M.S. Thesis*, Dept. Computer science, Air University, Islamabad, Pakistan, 2020.

[51] M. Gochoo, S. B. U. D. Tahir, A. Jalal and K. Kim, "Monitoring real-time personal locomotion behaviors over smart indoor-outdoor environments via body-worn sensors," *IEEE Access*, vol. 9, pp. 70556–70570, 2021.

[52] A. Jalal, N. Sarif, J. T. Kim and T. S. Kim, "Human activity recognition via recognized body parts of human depth silhouettes for residents monitoring services at smart home," *Indoor and Built Environment*, vol. 22, pp. 271–279, 2013.

[53] A. Arif and A. Jalal, "Automated body parts estimation and detection using salient maps and Gaussian matrix model," in *Proc of Int. Bhurban Conf. on Applied Sciences and Technologies (IBCAST)*, Islamabad, Pakistan, pp. 667–672, 2021.

[54] I. Akhter, A. Jalal and K. Kim, "Adaptive pose estimation for gait event detection using context-aware model and hierarchical optimization," *Journal of Electrical Engineering & Technology*, vol. 9, pp. 1–9, 2021.

[55] A. Jalal, A. Ahmed, A. A. Rafique and K. Kim, "Scene semantic recognition based on modified fuzzy c-mean and maximum entropy using object-to-object relations," *IEEE Access*, vol. 9, pp. 27758–27772, 2021.

[56] A. Jalal, M. Uddin and T. S. Kim, "Depth video-based human activity recognition system using translation and scaling invariant features for life logging at smart home," *IEEE Transactions on Consumer Electronics*, vol. 58, pp. 863–871, 2012.

[57] M. Gochoo, I. Akhter, A. Jalal and K. Kim, "Stochastic remote sensing event classification over adaptive posture estimation via multifused data and deep belief network," *Remote Sensing*, vol. 13, no. 5, pp. 1–29, 2021.

[58] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, no. 1, pp. 59–69, 1982.

[59] S. Kamal, A. Jalal and D. Kim, "Depth images-based human detection, tracking and activity recognition using spatiotemporal features and modified HMM," *Journal of Electrical Engineering & Technology*, vol. 11, no. 6, pp. 1857–1862, 2016.

[60] A. Jalal, S. Kamal and D. Kim, "Human depth sensors-based activity recognition using spatiotemporal features and hidden markov model for smart environments," *Journal of Computer Networks and Communications*, vol. 10, pp. 1–12, 2016.

[61] A. Jalal, S. Kamal and D. Kim, "Shape and motion features approach for activity tracking and recognition from kinect video camera," in *Proc of IEEE 29th Int. Conf. on Advanced Information Networking and Applications Workshops*, Gwangju, South Korea, pp. 445–450, 2015.

[62] A. Jalal, S. Kamal and D. Kim, "Facial expression recognition using 1D transform features and hidden markov model," *Journal of Electrical Engineering & Technology*, vol. 12, no. 4, pp. 1657–1662, 2017.

[63] S. Kaski, J. Nikkilä and T. Kohonen, "Methods for interpreting a self-organized map in data analysis," in *Proc of 6th European Symposium on Artificial Neural Networks (ESANN98)*, D-Facto, Brugfes, pp. 1–6, 1998.

[64] A. Jalal, Y. Kim, S. Kamal, A. Farooq and D. Kim, "Human daily activity recognition with joints plus body features representation using kinect sensor," in *Proc of Int. Conf. on Informatics, Electronics & Vision (ICIEV)*, Fukuoka, Japan, pp. 1–6, 2015.

[65] A. Jalal, S. Kamal, A. Farooq and D. Kim, "A spatiotemporal motion variation features extraction approach for human tracking and pose-based action recognition," in *Proc of Int. Conf. on Informatics, Electronics & Vision (ICIEV)*, Fukuoka, Japan, pp. 1–6, 2015.

[66] A. Jalal, S. Kamal and D. Kim, "Depth silhouettes context: A new robust feature for human tracking and activity recognition based on embedded HMMs," in *Proc of 12th Int. Conf. on Ubiquitous Robots and Ambient Intelligence (URAI)*, Goyang, South Korea, pp. 294–299, 2015.

[67] A. Jalal, S. Kamal and D. Kim, "Individual detection-tracking-recognition using depth activity images," in *Proc of 12th Int. Conf. on Ubiquitous Robots and Ambient Intelligence (URAI)*, Goyang, South Korea, pp. 450–455, 2015.

[68] A. Jalal, S. Kamal and C. A. Azurdia-Meza, "Depth maps-based human segmentation and action recognition using full-body plus body color cues via recognizer engine," *Journal of Electrical Engineering & Technology*, vol. 14, no. 1, pp. 455–461, 2019.

[69] A. Jalal, M. A. K. Quaid and M. A. Sidduqi, "A triaxial acceleration-based human motion detection for ambient smart home system," in *Proc of 16th Int. Bhurban Conf. on Applied Sciences and Technology (IBCAST)*, Islamabad, Pakistan, pp. 353–358, 2019.

[70] A. Jalal, M. Mahmood and A. S. Hasan, "Multi-features descriptors for human activity tracking and recognition in indoor-outdoor environments," in *Proc of Int. Bhurban Conf. on Applied Sciences and Technologies (IBCAST)*, Islamabad, Pakistan, pp. 371–376, 2019.

[71] A. Jalal, A. Nadeem and S. Bobasu, "Human body parts estimation and detection for physical sports movements," in *Proc of 2nd Int. Conf. on Communication, Computing and Digital Systems (C-CODE)*, Islamabad, Pakistan, pp. 104–109, 2019.

[72] A. Ahmed, A. Jalal and K. Kim, "RGB-D images for object segmentation, localization and recognition in indoor scenes using feature descriptor and hough voting," in *Proc of 17th Int. Bhurban Conf. on Applied Sciences and Technology (IBCAST)*, Islamabad, Pakistan, pp. 290–295, 2020.

[73] M. Mahmood, A. Jalal and K. Kim, "White stag model: Wise human interaction tracking and estimation (WHITE) using spatio-temporal and angular-geometric (STAG) descriptors," *Multimedia Tools and Applications*, vol. 79, pp. 6919–6950, 2020.

[74] M. A. K. Quaid and A. Jalal, "Wearable sensors based human behavioral pattern recognition using statistical features and reweighted genetic algorithm," *Multimedia Tools and Applications*, vol. 79, pp. 6061–6083, 2020.

[75] A. A. Rafique, A. Jalal and A. Ahmed, "Scene understanding and recognition: statistical segmented model using geometrical features and Gaussian naïve Bayes," in *Proc of IEEE Conf. on Int. Conf. on Applied and Engineering Mathematics*, Islamabad, Pakistan, vol. 57 pp. 1–6, 2019.

[76] A. Ahmed, A. Jalal and A. A. Rafique, "Salient segmentation based object detection and recognition using hybrid genetic transform," in *Proc of Int. Conf. on Applied and Engineering Mathematics (ICAEM)*, Islamabad, Pakistan, pp. 203–208, 2019.

[77] A. A. Rafique, A. Jalal and K. Kim, "Statistical multi-objects segmentation for indoor/outdoor scene detection and classification via depth images," in *Proc of 17th Int. Bhurban Conf. on Applied Sciences and Technology (IBCAST)*, Islamabad, Pakistan, pp. 271–276, 2020.

[78] A. Jalal, M. A. K. Quaid and K. Kim, "A wrist worn acceleration based human motion analysis and classification for ambient smart home system," *Journal of Electrical Engineering & Technology*, vol. 14, no. 4, pp. 1733–1739, 2019.

[79] A. Nadeem, A. Jalal and K. Kim, "Human actions tracking and recognition based on body parts detection via artificial neural network," in *Proc of 3rd Int. Conf. on Advancements in Computational Sciences (ICACS)*, Lahore, Pakistan, pp. 1–6, 2020.

[80] S. A. Rizwan, A. Jalal and K. Kim, "An accurate facial expression detector using multi-landmarks selection and local transform features," in *Proc of 3rd Int. Conf. on Advancements in Computational Sciences (ICACS)*, Lahore, Pakistan, pp. 1–6, 2020.

[81] K. Kim, A. Jalal and M. Mahmood, "Vision-based human activity recognition system using depth silhouettes: A smart home system for monitoring the residents," *Journal of Electrical Engineering & Technology*, vol. 14, pp. 2567–2573, 2019.

[82] A. Ahmed, A. Jalal and K. Kim, "Region and decision tree-based segmentations for multi-objects detection and classification in outdoor scenes," in *Proc of Int. Conf. on Frontiers of Information Technology (FIT)*, Islamabad, Pakistan, pp. 209–214, 2019.

[83] A. Ahmed, A. Jalal and K. Kim, "A novel statistical method for scene classification based on multi-object categorization and logistic regression," *Sensors*, vol. 20, pp. 3871, 2020.

[84] A. Jalal, N. Khalid and K. Kim, "Automatic recognition of human interaction via hybrid descriptors and maximum entropy markov model using depth sensors," *Entropy*, vol. 22, pp. 817, 2020.

[85] M. Batool, A. Jalal and K. Kim, "Telemonitoring of daily activity using accelerometer and gyroscope in smart home environments," *Journal of Electrical Engineering & Technology*, vol. 15, pp. 1–9, 2020.

[86] A. Jalal, M. Batool and K. Kim, "Stochastic recognition of physical activity and healthcare using tri-axial inertial wearable sensors," *Applied Sciences*, vol. 10, no. 20, pp. 7122, 2020.

[87] A. Jalal, M. A. K. Quaid, S. B. Tahir and K. Kim, "A study of accelerometer and gyroscope measurements in physical life-log activities detection systems," *Sensors*, vol. 20, no. 22, pp. 6670, 2020.

[88] A. A. Rafique, A. Jalal and K. Kim, "Automated sustainable multi-object segmentation and recognition via modified sampling consensus and kernel sliding perceptron," *Symmetry*, vol. 12, no. 11, pp. 1928, 2020.

[89] M. Javeed, A. Jalal and K. Kim, "Wearable sensors based exertion recognition using statistical features and random forest for physical healthcare monitoring," in *Proc of 17th Int. Bhurban Conf. on Applied Sciences and Technology (IBCAST)*, Islamabad, Pakistan, pp. 2512–517, 2020.

[90] M. D. Rodriguez, J. Ahmed and M. Shah, "Action mach a spatio-temporal maximum average correlation height filter for action recognition," in *Proc of IEEE Conf. on Computer Vision and Pattern Recognition*, Anchorage, USA, pp. 1–8, 2008.

[91] S. Safdarnejad, X. Liu, L. Udpa, B. Andrus, J. Wood *et al.,* "Sports videos in the wild (SVW): A video dataset for sports analysis," in *Proc of 11th IEEE Int. Conf. and Workshops on Automatic Face and Gesture Recognition (FG)*, Ljubljana, Slovenia, pp. 1–7, 2015.

[92] A. Jalal, M. Batool and K. Kim, "Sustainable wearable system: Human behavior modeling for life-logging activities using K-ary tree hashing classifier," *Sustainability*, vol. 12, no. 24, pp. 10324, 2020.

[93] U. Azmat and A. Jalal, "Smartphone inertial sensors for human locomotion activity recognition based on template matching and codebook generation," in *Proc of Int. Conf. on Communication Technologies*, Tianjin, China, pp. 109–114, 2021.

[94] A. Ahmed, A. Jalal and K. Kim, "A novel statistical method for scene classification based on multi-object categorization and logistic regression," *Sensors*, vol. 20, no. 14, pp. 3871, 2020.

[95] N. Khalid, Y. Y. Ghadi, M. Gochoo, A. Jalal and K. Kim, "Semantic recognition of human-object interactions via Gaussian-based elliptical modelling and pixel-level labeling," *IEEE Access*, vol. 9, pp. 111249–111266, 2021.

[96] S. Park and J. K. Aggarwal, "Segmentation and tracking of interacting human body parts under occlusion and shadowing," in *Proc of Workshop on Motion and Video Computing*, Florida, USA, pp. 105–111, 2002.

[97] S. Li and A. B. Chan, "3D human pose estimation from monocular images with deep convolutional neural network," in *Proc of Asian Conf. on Computer Vision*, Singapore, pp. 332–347, 2014.

[98] H. -W. Chen and M. McGurr, "Moving human full body and body parts detection, tracking and applications on human activity estimation, walking pattern and face recognition," *Automatic Target Recognition XXVI*, vol. 984, pp. 984401T1–34, 2016.

[99] C. Rodriguez, B. Fernando and H. Li, "Action anticipation by predicting future dynamic images," in *Proc of the European Conf. on Computer Vision (ECCV)*, Munich, Germany, pp. 1–16, 2018.

[100] D. Xing, X. Wang and H. Lu, "Action recognition using hybrid feature descriptor and VLAD video encoding," in" *Proc of Asian Conf. on Computer Vision*, Singapore, pp. 99–112, 2014.

[101] C. Chattopadhyay and S. Das, "Supervised framework for automatic recognition and retrieval of interaction: A framework for classification and retrieving videos with similar human interactions," *IET Computer Vision*, vol. 10, no. 3, pp. 220–227, 2016.

[102] S. Sun, Z. Kuang, L. Sheng, W. Ouyang and W. Zhang, “Optical flow guided feature: a fast and robust motion representation for video action recognition,” in *Proc of the IEEE Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, USA, pp. 1390–1399, 2018.

[103] R. F. Rachmadi, K. Uchimura and G. Koutaki, “Combined convolutional neural network for event recognition,” in *Proc of the Korea-Japan Joint Workshop on Frontiers of Computer Vision*, Mokpo, South Korea, pp. 85–90, 2016.

[104] Y. Zhu, K. Zhou, M. Wang, Y. Zhao and Z. Zhao, “A comprehensive solution for detecting events in complex surveillance videos,” *Multimedia Tools and Applications*, vol. 78, no. 1, pp. 817–838, 2019.