remote sensing

MDPI

Article

Stochastic Remote Sensing Event Classification over Adaptive Posture Estimation via Multifused Data and Deep Belief Network

Munkhjargal Gochoo [1], Israr Akhter [2], Ahmad Jalal [2] and Kibum Kim [3,*]

[1] Department of Computer Science and Software Engieering, United Arab Emirates University, Al Ain 15551, United Arab Emirates; mgochoo@uaeu.ac.ae
[2] Department of Computer Science, Air University, Islamabad 44000, Pakistan; israrakhter.edu@gmail.com (I.A.); ahmadjalal@mail.au.edu.pk (A.J.)
[3] Department of Human-Computer Interaction, Hanyang University, Ansan 15588, Korea
* Correspondence: kibum@hanyang.ac.kr

Abstract: Advances in video capturing devices enable adaptive posture estimation (APE) and event classification of multiple human-based videos for smart systems. Accurate event classification and adaptive posture estimation are still challenging domains, although researchers work hard to find solutions. In this research article, we propose a novel method to classify stochastic remote sensing events and to perform adaptive posture estimation. We performed human silhouette extraction using the Gaussian Mixture Model (GMM) and saliency map. After that, we performed human body part detection and used a unified pseudo-2D stick model for adaptive posture estimation. Multifused data that include energy, 3D Cartesian view, angular geometric, skeleton zigzag and moveable body parts were applied. Using a charged system search, we optimized our feature vector and deep belief network. We classified complex events, which were performed over sports videos in the wild (SVW), Olympic sports, UCF aerial action dataset and UT-interaction datasets. The mean accuracy of human body part detection was 83.57% over the UT-interaction, 83.00% for the Olympic sports and 83.78% for the SVW dataset. The mean event classification accuracy was 91.67% over the UT-interaction, 92.50% for Olympic sports and 89.47% for SVW dataset. These results are superior compared to existing state-of-the-art methods.

Keywords: deep belief network; event classification; human body part detection; multifused data; pseudo-2D-stick model

Citation: Gochoo, M.; Akhter, I.; Jalal, A.; Kim, K. Stochastic Remote Sensing Event Classification over Adaptive Posture Estimation via Multifused Data and Deep Belief Network. *Remote Sens.* **2021**, *13*, 912. https://doi.org/10.3390/rs13050912

Academic Editors: Józef Lisowski, Jorge Garcia-Gutierrez and Maria del Mar Martinez-Ballesteros

Received: 14 January 2021
Accepted: 23 February 2021
Published: 28 February 2021

1. Introduction

The digital era, data visualization and advancements in technology enable us to analyze digital data, human-based images and videos and multimedia contents [1–3]. Due to globalization and the convenience of data transmission, it is now possible and important to examine multimedia data for surveillance; emergency services; educational institutions; national institutions, such as law enforcement; and the activities of various people from employees to criminals. National databases with records of citizens, hospitals, monitoring systems, traffic control systems and factory observation systems are just a few examples of multimedia-based contents [4–9]. The developments of Adaptive Posture Estimate Systems (APES) and Event Classification Methods (ECM) are hot topics and challenging domains in recent decades. A large amount of progress has been made by researchers who are innovating advanced frameworks, but there remain many challenges [10–13]. Event classification and adaptive posture estimation are used in many applications, such as airport security systems, railways, bus stations and seaports, where normal and abnormal events can be detected in real-time [14–18]. Sports events can be classified using Adaptive Posture Estimation Systems (APES) and Event Classification Methods (ECM) mechanisms whether the events occur indoors or outdoors [19]. Adaptive Posture Estimation Systems

(APES) and Event Classification Methods (ECM) open new doors in technology and applied sciences domains to save manpower, time and costs, and to make prudent decisions at the right times [20]. Adaptive Posture Estimation Systems (APES) and Event Classification Methods (ECM) still need to be improved in order to accurately extract features from videos and images, and to estimate and track human motion, human joints movement and event classification.

In this paper, we propose a unified framework for stochastic remote sensing event classification and adaptive posture estimation. A pseudo-2D stick mesh model is implemented via a Multifused Data extraction approach. These features extract various optimal values, including energy, skeleton zigzag, angular geometric, 3D Cartesian and moveable body parts. For data optimization, we used the meta-heuristic charged system search (CSS) algorithm, and event classification was performed by the deep belief network (DBN). The main contributions of this research paper are as follows:

- We contribute a robust method for the detection of nineteen human body parts over complex human movement; challenging events and human postures can be detected and estimated more accurately.
- For more accurate results in adaptive posture estimation and classification, we designed a skeletal pseudo-2D stick model that enables the detection of nineteen human body parts.
- In the multifused data, we extracted sense-aware features which include energy, moveable body parts, skeleton zigzag features, angular geometric features and 3D Cartesian features. Using these extracted features, we can classify stochastic remote sensing events in multiple human-based videos more accurately.
- For data optimization, a hierarchical optimization model is implemented to reduce computational cost and to optimize data, and charged system search optimization is implemented over-extracted features. A deep belief network is applied for multiple human-based video stochastic remote sensing event classification.

The plan of this research article is as follows: Section 2 contains a detailed overview of related works. In Section 3, the methodology of Adaptive Posture Estimation and Event Classification (APEEC) is discussed. Section 4 describes the complete description of the experimental setup and a comprehensive comparison of the proposed system with existing state-of-the-art systems. In Section 5, future directions and conclusions are defined.

2. Related Work

Advances in camera technologies, video recording and body marker sensor-based devices enable superior approaches to the farming and analysis of information for research and development in this field. The research community has contributed many novel, robust, and innovative methods to identify human events, actions, activities and postures. Table 1 contains a detailed overview of the related work.

Table 1. Related work and main contributions.

Methods	Main Contributions
Lee et al. [21]	They developed a state-of-the-art hierarchical method in which human body part identification is used for critical silhouette monitoring. Additionally, they introduced region comparison features for optimal data values and to obtain rich information.
Aggarwal et al. [22]	They designed a robust scheme, for human body part motion analysis, using multiple cameras that track the human body parts, and for human identification. They also developed a 2D–3D projection for human body joints.
Wang et al. [23]	They explained a framework to analyze human behavior. For this, they used the identification of humans, human activity identification and human tracking approaches.

Table 1. *Cont.*

Methods	Main Contributions
Liu, J et al. [24]	They proposed a robust random forest based technique for human body part training, using temporal features, static and motion features. They estimated various human actions in videos and images.
Khan and M. A [25]	They suggested an automated process using multiview features, vertical and horizontal gradient features. They used Deep Neural Network (DNN) fusion to identify human actions. A pre-trained Convolutional Neural Network CNN-VGG19 model is adopted to achieve DNN-based feature techniques.
Zou and Shi [26]	They explained an automated system, Adaptation-Oriented Features (AOF), with one-shot action identification for the estimation of human actions. The system pertains to every class, and for output, they combine AOF parameters.
Franco and Magnani [27]	They designed a multilayer structure for intensive human skeleton information, via RGB images. They extracted Histogram of Oriented (HOG) descriptor features for human action recognition.
Ullah and Muhammad [28]	They describe a unified Convolutional Neural Network (CNN)-based approach for real-time data communication and data streams. For data extraction from non-monitoring devices, they used visual sensors. To monitor the human action, temporal features along with deep autoencoder and deep features are estimated via the Convolutional Neural Network (CNN) model.
Jalal, A et al. [29]	They proposed a novel technique to identify daily human activities in a smart home environment via depth-based daily routine functionality. They also defined the AR method for the estimation of human activities.
Jalal, A et al. [30]	They explained a robust system using a marker-based system in which many body markers are attached to the human body at various strategic points. With the help of extracted synthetic contexts and the multiactors platform, they identify human activity.
Kruk and Reijne [31]	They developed a unified model to estimate vibrant human motion in sports event via body marker sensors. The main contribution is the estimation of the kinematics of human body joints, acceleration, velocity and the reconstruction of the human pose to compute human events in sports datasets.
Wang and Mori [32]	They proposed a novel technique for event recognition, via spatial relations and human body pose. Tree-based features are described using the kinematics information of connected human body parts.
Amft and Troster [33]	They developed a robust framework via a Hidden Markov approach. Time-continuous based features using body marker sensors event recognition is achieved.
Wang et al. [34]	They designed a new systematic approach to estimate the consistency of human motion with the help of a human tracking approach. A Deep Neural Network (DNN) is used for event recognition.
Jaing et al. [35]	They introduced a multilayered feature method for the estimation of human motion and movements. For event recognition in dynamic scenes, they used a late average mixture algorithm.
Li, et al. [36]	They proposed an innovative method for event recognition via joint optimization, optical flow and a histogram of the obtained optical flow. With the help of the norm optimization method, body joint reconstruction and a Low-rank and Compact Coefficient Dictionary Learning (LRCCDL) approach, they achieved accurate event identification.
Einfalt et al. [37]	They designed a unified method for the event recognition of an athlete in motion, using task classification, extraction of chronological 2D posture features and a convolutional sequence network. They recognized a sports event precisely.
Yu et al. [38]	They explain a heuristic framework that can detect events in a distinct interchange from soccer competition videos. This is achieved with the help of the replay identification approac to discover maximum context features for gratifying spectator requirements and constructing replay story clips.
Franklin et al. [39]	They proposed a robust deep learning mechanism for abnormal and normal event detection. Segmentation, classification and graph-based approaches were used to obtain the results. Using deep learning methods, they found normal and abnormal features for event interval utilization.
Lohithashva et al. [40]	They designed an innovative mixture features descriptor approach for intense event recognition via the Gray Level Co-occurrence Matrix (GLCM) and the Local Binary Pattern (LBP). They used extracted features with machine learning supervised classification systems for event identification.
Feng et al. [41]	They proposed a directed Long Short Term Memory (LSTM) method using a Convolutional Neural Network (CCN)-based model to extract deep features' temporal positions in composite videos. The state-of-the-art YOLO v3 model is used for human identification and a guided Long Short Term Memory (LSTM)-based framework is adopted for event recognition.

Table 1. *Cont.*

Methods	Main Contributions
Khan et al. [42]	They developed a body-marker sensor-based technique for home-based patient management. Body-marker sensors utilizing a color indication scheme are attached to the joints of the human body to record data of the patients.
Esfahani et al. [43]	For sports events, human motion observation body-marker instruments were used to develop a low computational process-based Trunk Motion Method (TMM) with Body-worn Sensors (BWS). In this approach, 12 removable sensors were utilized to calculate trunk 3D motions.
Golestani et al. [44]	They proposed a robust wireless framework to identify physical human actions. They tracked human actions with a magnetic induction cable; body-marker sensors were associated with human body joints. To achieve improved accuracy, the laboratory estimation function and Deep Recurrent Neural Network (RNN) were used.

3. Proposed System Methodology

RGB video-based cameras are utilized to record video data as input of the proposed system during preprocessing; frame conversion and noise removal are applied after which human silhouette extraction, human detection and human body part identification via the 2D stick model are performed. After this, the pseudo-2D stick model is evaluated for human posture estimation, and multifused data are used for feature vector extraction. A charged system search (CSS) [45] algorithm is used for optimization and event classification. We used a machine learning model named the deep belief network (DBN) [46]. Figure 1 illustrates the proposed Adaptive Posture Estimation and stochastic remote sensing Event Classification (APEEC) system architecture.

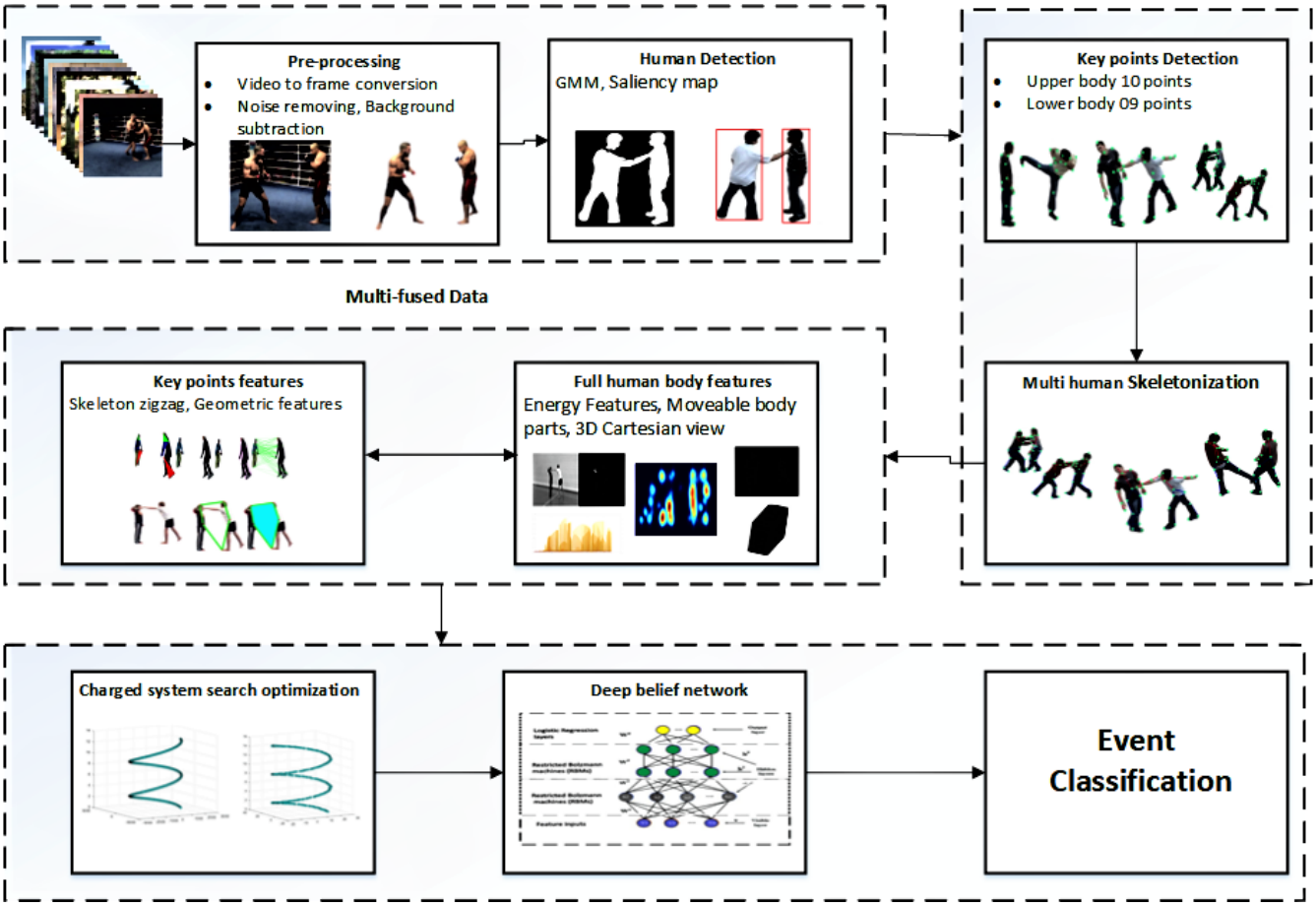


Figure 1. The system architecture of the proposed Posture Estimation and Event Classification (APEEC) system.

3.1. Preprocessing Stage

Data preprocessing is among the main steps which is adopted to avoid extra data processing cost. In the preprocessing step, video to image conversion was performed; then, grayscale conversion, using Gaussian filter noise removal techniques were used to minimize superfluous information. After that, using the change detection technique and Gaussian Mixture Model (GMM) [47], we performed initial background subtraction for further processing. Then, to extract the human silhouette, the saliency map technique [48] was adopted in which saliency values were estimated. Saliency SV for the pixel $(i,\ j)$ was calculated as

$$\mathrm{SV}(x,y) \ = \ \sum_{(m,n)\in N} d[V(x,y), Q(m,n)] \tag{1}$$

where N is denoted as the region near to the saliency pixel at $(x,\ y)$ position and d represents the locus difference among pixel vectors V and Q. After the estimation of saliency values for all the certain areas of the input image, a heuristic threshold technique was used to distinguish the foreground from the background. Figure 2 shows the results of the background subtraction, human silhouette extraction and the results of human detection.

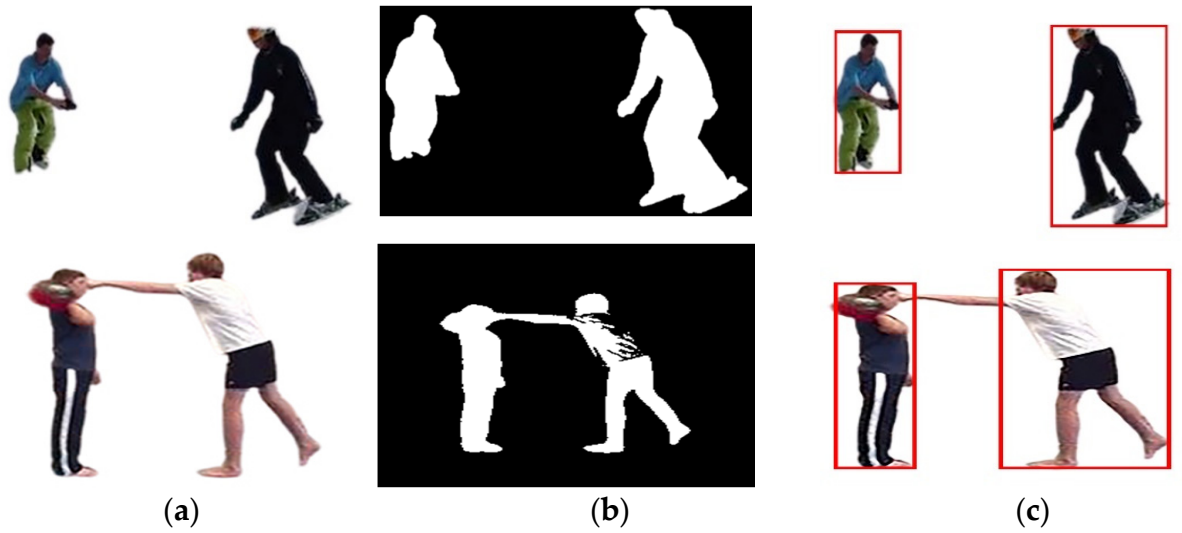


Figure 2. Preprocessing steps for silhouette extraction. (**a**) Background subtracted (**b**) human silhouette extraction and (**c**) human defection in RGB images and videos.

Figure 3 shows the results of multiperson tracking and detection for the UCF aerial action dataset.

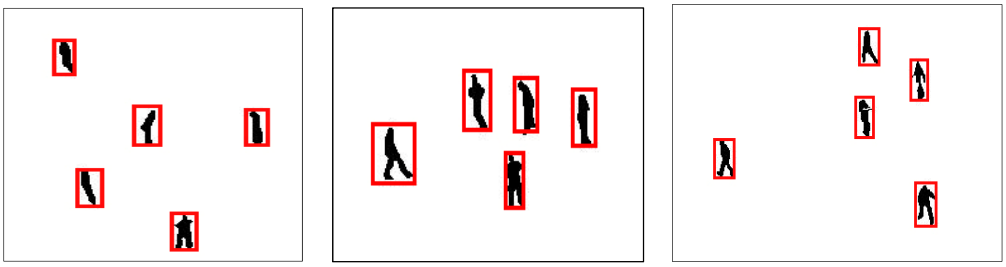


Figure 3. A few examples of multiperson tracking and detection for the UCF aerial action dataset.

After successfully extracting the human silhouette, detection of human body parts was performed.

3.2. Posture Estimation: Body Part Detection

During human posture estimation and identification of human body points, the estimation of detected human silhouette's outer shape values Hpv was used to estimate the center torso point. The recognition of the human torso point is expressed as

$$K_{Top}^{f} \leftarrow K_{To}^{f-1} + \Delta K_{To}^{f-1} \tag{2}$$

where K_{Top}^{f} denotes a human torso point position in any given frame f, which is the result of computing by the frame variances. For the recognition of the human ankle position, we considered the point 1/4 between the foot and the knee points. Equation (3) shows the human ankle point

$$K_{SA}^{f} = (K_{SF}^{f} - K_{SK}^{f})/4 \tag{3}$$

where K_{SA}^{f} is the ankle position, K_{SF}^{f} is the foot position and K_{SK}^{f} represents the human knee point. For wrist point estimation, we considered the point 1/4 of the value of the distance between the hand and elbow points, which is represented in Equation (4) as

$$K_{SW}^{f} = (K_{SHN}^{f} - K_{SEL}^{f})/4 \tag{4}$$

where K_{SW}^{f} is the human wrist point, K_{SHN}^{f} is the hand point and K_{SEL}^{f} represents the elbow point. Algorithm 1 shows the detailed description of human body part detection.

Algorithm 1. Human body key point detection.

Input: H_{ES}: Extracted full human silhouette
Output: 19 body parts as head, shoulders, neck, wrists, elbows, hands, torso, hips, ankle, knees and foot. H_{FS} = Full human silhouette, H_{Sh} = human shape, H_s = height, W_s = width, L_s = left, R_s = right, I_{sh} = head, I_{sn} = neck

Repeat
For i = 1 to N **do**
Search (H_{FS})
Ish = Human_head;
S_Up = UpperPoint (I_{sh})
S_ehp = Human _End_Head_point (I_{sh})
S_Mp = mid (H_s,W_s)/2
S_Fp = Bottom(HFS)& search(L_s, R_s)
S_Knp = mid(S_Mp, S_ Fp)
S_Hnp = S_ehp & S_Up & search(L_s, R_s)
S_Shp = search(I_{sh}, I_{sh}) *& search*(R_s, L_s)
S_Elp = mid(S_Hnp, S_Shp)
S_Wrp = mid (S_Hnp, S_Elp)/2
S_Hip = S_Mp &search(R_s,L_s)
S_Anp = mid (S_Knp, S_Fp)/4
End
Until complete human silhouette searched.
return 19 body parts as head, shoulders, neck, wrists, elbows, hands, torso, hips, ankles, knees and foot. H_{FS} = Full human silhouette, H_{Sh} = human shape, H_s = height, W_s = width, L_s = left, R_s = right, I_{sh} = head, I_{sn} = neck

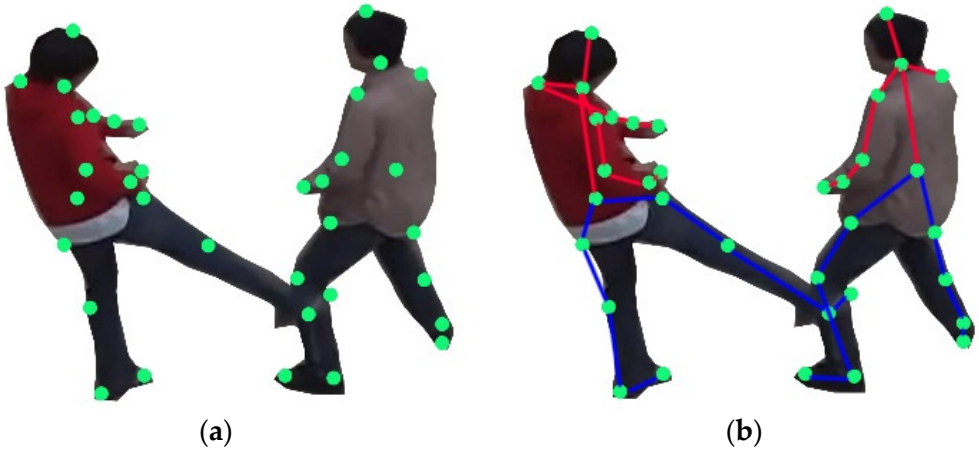
In this segment, the human skeletonization over-extracted body points [48,49] are denoted as a pre-pseudo-2D stick approach. Figure 4 shows the comprehensive overview of the pre-pseudo-2D stick model that includes 19 human body points, which are considered as three key skeleton fragments: human upper body segment ($HUbs$), human midpoint segment (HMp) and human lower body segment ($HLbs$). $HUbs$ is based on the linkage of the head (Ish), neck (Isn), shoulders (S_Shp), elbow (S_Elp), wrist (S_Wrp) and hand points (S_Hnp). $HLbs$ is founded via the association of hips (S_Hip), knees (S_Hnp), ankle

(S_Anp) and foot (S_Fp). Each body point takes a particular time t to accomplish a specific action. Equations (5)–(7) show the mathematical relationships of the pre-pseudo-2D stick model as

$$HUbs = Ish \bowtie Isn \bowtie S\_Shp \bowtie S\_Elp \bowtie S\_Wrp \bowtie S\_Hnp \quad (5)$$

$$HMp = S\_Mp \bowtie HUbs \quad (6)$$

$$HLbs = S\_Hip \bowtie S\_Hnp \bowtie S\_Anp \bowtie S\_Fp \bowtie HMp \quad (7)$$



(**a**) (**b**)

Figure 4. Human body point detection and 2D stick model. (**a**) Nineteen human body points are detected; (**b**) 2D stick model over nineteen body points.

Due to the massive distance from the drone to object in UCF aerial action dataset, it was difficult to find accurate human body parts. Figure 5 represents the body point results over the UCF aerial action dataset.

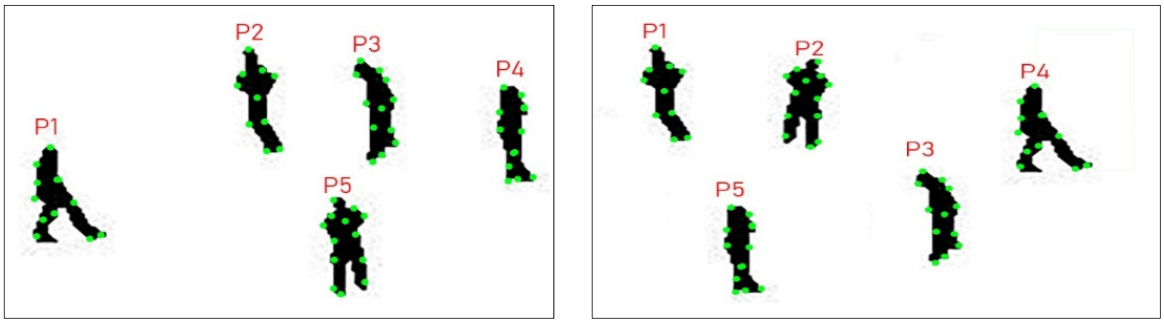


Figure 5. Human body point detection over the UCF aerial action dataset.

3.3. Posture Estimation: Pseudo-2D Stick Model

In this segment, we proposed a pseudo-2D stick approach that empowers an indestructible human skeleton throughout human motion [49]. To perform this, we identified nineteen human body points, after which interconnection processing for every node was performed using a self-connection technique [50]. Then, the 2D stick model (Section 3.2) was applied based on the concept of fixed undirected skeleton mesh. For lower and upper movements, we used stick scaling; 15 pixels is the threshold limit of stick scaling; if this is exceeded, the fixed undirected skeleton mesh will not accomplish the required results. Equation (8) represents the mathematical formulation of the human body stick scaling.

$$Sm_{bS} = Hps\left\{ \begin{array}{c} 1,\ if\ Up \,||\, Sm \leq 20 \\ 0,\ Up \,||\, Sm > 20 \end{array} \right\} \quad (8)$$

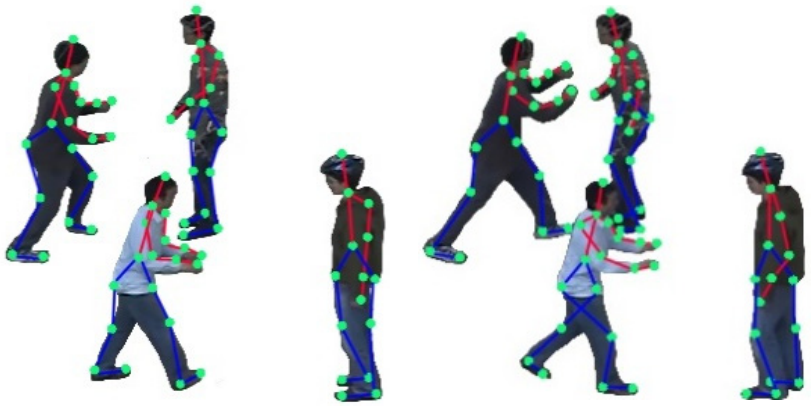
where Sm_{bS} symbolizes as a human fixed 2D-stick mesh, Up denotes the upper limit, L is for the lower limit of stick scaling and Hps denotes the human body part scaling. To track the human body parts, we allowed the human skeleton to use kinematic and volumetric data. The size of the human outer shape was used to calculate the lower and upper distances of the human silhouette. After that, the measurement of the given frame was estimated using the given frame size. Equation (9) calculates the procedure for identifying the head location.

$$K_{Ish}^{f} \leftarrow K_{Ish}^{f-1} + \Delta K_{Ish}^{f-1} \tag{9}$$

while K_{Ish}^{f} represents the head location in any given frame. Human body movement direction change recognition, which arose in frame 1 to the next frame, was used as the pre-step of pseudo-2D. To perform the complete pseudo-2D stick model, the degree of freedom and the edge information of the human body were used; global and local coordinate methods were implemented, which helped us determine the angular movements of human body parts. While the global and local coordinate methods were performed, to achieve the final results of the pseudo-2D stick model, we implement the Cartesian product [21]. Figure 6 shows a few example results of the pseudo-2D stick model, and Algorithm 2 represents the complete overview of the pseudo-2D stick model.

Algorithm 2. Pseudo-2D stick model.

Input: Human body key point and 2D stick model
Output: Pseudo-2D stick graph (kp_1, kp_2, kp_3, . . . , kp_n)
Dkp = Detection of human body key points, Cdp = Connection of detected points, Ss = Sticks scaling, Sg = 2D skeleton graph, Vi = Volumetric information, Kk = kinematic dependency and key points tracing, Ei = edges information, Dof = Degree of freedom, GL = Global and local coordinates, Sc = Skeleton Graph Cartesian product
% initiating Pseudo-2D Stick Graph %
Pseudo-2D Stick Graph ← []
P2DSG_Size ← Get P2DSG_Size ()
% loop on extracted human silhouettes%
For I = 1:K
P2DSG_interactions ← GetP2DSG_interactions
%Extracting Dkp, Cdp, Ss, Sg, Vi, Kk, Ei, Dof, GL, Sc%
Detection of body key points ← Dkp(P2DSG_interactions)
Connection of detected points ← Cdp(P2DSG_interactions)
Sticks scaling ← Ss(P2DSG_interactions)
2D skeleton graph ← Sg(P2DSG_interactions)
Volumetric information ← Vi(P2DSG_interactions)
Kinematics dependency ← Kk(P2DSG_interactions)
Edges information ← Ei(P2DSG_interactions)
Degree of freedom ← Dof(P2DSG_interactions)
Global and Local coordinates ← GL(P2DSG_interactions)
Skeleton Graph Cartesian product ← Sc(P2DSG_interactions)
Pseudo-2D Stick Graph ← Get P2DSG
Pseudo-2D Stick Graph.append (P2DSG)
End
Pseudo-2D Stick Graph ← Normalize (Pseudo-2D Stick Graph)
return Pseudo-2D Stick Graph (kp_1, kp_2, kp_3, . . . , kp_n)



(**a**) (**b**)

Figure 6. Pseudo-2D stick model. (**a**) Push event first state; (**b**) push event second state.

3.4. Multifused Data

In this segment, we give a comprehensive overview of multifused data, including skeleton zigzag, angular geometric, 3D Cartesian view, energy and moveable body part features for APEEC. Algorithm 3 describes the formula for multifused data extraction.

3.4.1. Skeleton Zigzag Feature

In skeleton zigzag features, we defined human skeleton points as human outer body parts. Initially, we calculated skeleton zigzag features via the Euclidean distance in-between body parts of the first human silhouette and those of the second silhouette. This distance vector helped us to find more accurate stochastic remote sensing event classification and human posture estimation. Using Equation (10), we determined the outer distance between two human silhouettes. Figure 7 represents the skeleton zigzag features results.

$$Szf \quad = \quad h1\_dis(f_1.f_2) \leftrightarrow h2\_dis(f_1.f_1) \tag{10}$$

where Szf is the skeleton zigzag features, $h1\_dis$ is the distance of the first human silhouette and $h2\_dis$ is the distance of the second human silhouette.

(**a**) (**b**)

Figure 7. Skeleton zigzag features (**a**) from ski class results over SVW dataset; (**b**) push event over UT-interaction dataset.

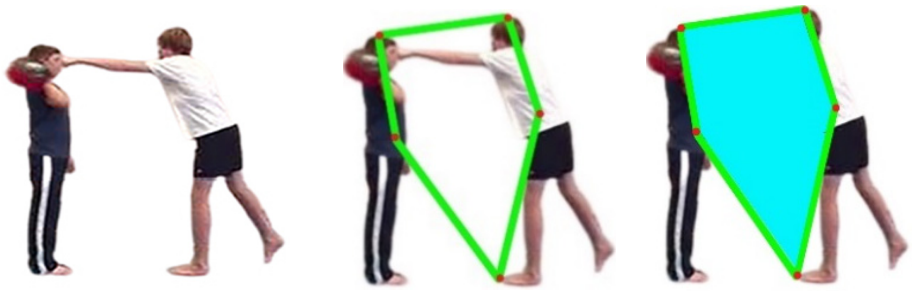
3.4.2. Angular Geometric Feature

In angular geometric features, we considered an orthogonal shape over human body parts. We considered five basic body parts as edges of the orthogonal body in which

head point, torso point and feet point were included. We drew an orthogonal shape and computed the area using Equation (11) and put the results in the main features vector.

$$Agf = \left(\frac{5}{2}\right) * s * a \tag{11}$$

where Agf is the angular geometric feature vector, $5/2$ is a constant, s is the side of a pentagon and a denotes the apothem length. Figure 8 shows the results of the angular geometric features.



(**a**) (**b**) (**c**)

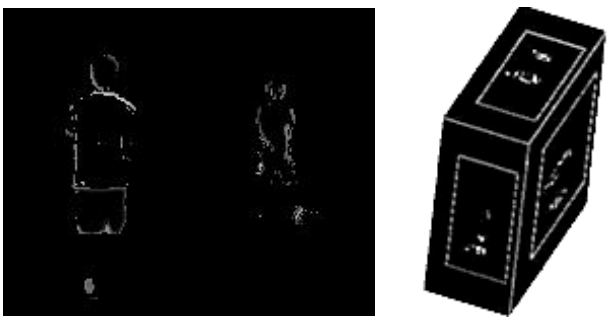
Figure 8. Example of angular geometric features results: (**a**) simple RGB image; (**b**) bounding box of angular geometric features; (**c**) the covered area region is considered to be the angular geometric feature.

3.4.3. 3D Cartesian View Feature

From multifused data, we determined the smoothing gradient from the extracted human silhouette and estimate the gradient indexes of the detected full human body silhouette. After this, we obtained a 3D Cartesian product and a 3D Cartesian view of the extracted smoothing gradient values. By this, we could obtain the 3D indexes. After that, the difference between every two consecutive frames f and $f-1$ of the human silhouettes H_{FS} was calculated. Equation (12) represents the mathematical formulation for the estimated 3D Cartesian view. After estimating the 3D values, we placed them in a trajectory and concatenated them with the central feature vector as

$$CVI_{TSV}(f) = \left| H_{FS}{}^{f}_{TSV} - H_{FS}{}^{f-1}_{TSV} \right| \tag{12}$$

where CVI represents the 3D Cartesian view vector; TSV denotes the side, front and top views of the extracted 3D Cartesian view. Figure 9 represents the results of the 3D Cartesian view and the 2D representation.



(**a**) (**b**)

Figure 9. Example of 3D Cartesian view over SVW dataset: (**a**) 2D representation; (**b**) 3D representation of 3D Cartesian view.

3.4.4. Energy Feature

In the energy feature, *Egn(t)* calculated the motion of the human body part in the energy-based matrix, which holds a set of energy values [0–10,000] over the identified human silhouette. After the circulation of energy value, we collected only the upper energy value using the heuristic thresholding technique and placed all extracted values in a 1D array. The mathematical representation of energy distribution is shown in Equation (13), and example results of energy features are represented in Figure 10.

$$Egn(t) = \sum_{0}^{w} IamgR(i) \tag{13}$$

where $Egn(t)$ specifies the energy array vector, i expresses index values and $IamgR$ represents the index value of certain RGB pixels.

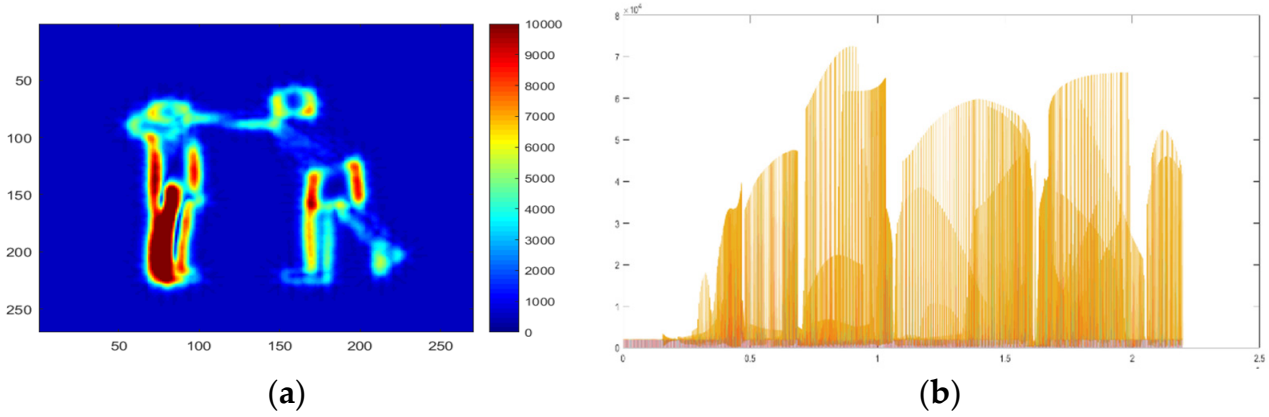


Figure 10. (**a**) Results of energy features over the SVW dataset; (**b**) 1D vector representation of energy features.

3.4.5. Moveable Body Parts Feature

In moveable body parts features, only relocated body parts of the human were considered. To identify these body parts features, we considered the moveable section of the human body parts in the preceding frame as the main spot, to crop the given frame patch p of size IxJ in the present frame and to estimate the output value as

$$\hat{S} = ft^{-1}(A \bigodot ft(n^{\hat{x}p})) \tag{14}$$

where ft^{-1} denotes the inverse Fourier transform, $\bigodot$ denotes the matrix Hadamard product, n is a correlation, $\hat{x}$ is the output value of the marked shape in a certain image and $\hat{S}$ shows the similarity among the candidate portion of the frame and the preceding region. Thus, the present location of the moveable body parts can be identified by obtaining the higher values of $\hat{S}$ as;

$$H = max(\hat{\hat{S}}) \tag{15}$$

However, when the transformed regions were recognized, we increased the bonding region across the moving body points, found the pixel's location and traced additional moveable body parts in the series of frames as

$$Mb = \sum_{0}^{Nk} MI(Nk) \tag{16}$$

where Mb is the moving body parts vector, Nk is the integer index and MF denotes the location of pixel values. Figure 11 describes the results of moveable body parts features. Algorithm 3 shows the detailed procedures of the feature extraction framework.

Algorithm 3. Multifused data.

Input: N: Extracted human silhouettes frames of RGB images
Output: sense –aware feature vectors(sf_1, sf_2, sf_3, . . . , sf_n)
% initiating feature vector for stochastic remote sensing event classification %
sense-awarefeature_vectors ← []
F_vectorsize ← GetVectorsize ()
% loop on extracted human silhouettes frames %
For i = 1:K
F_vectors_ interactions ← Get_F_vectors(interactions)
% extracting energy features, moveable human body parts, Smoothing gradient 3D Cartesian view, key points angle features, human skeleton features%
Energy Features ←ExtractEnergyFeatures(F_vectors_ interactions)
MoveableHumanBodyparts←ExtractMoveableHumanbodyparts(F_vectors_ interactions)
Smoothing Gradient 3D Cartesian View ← ExtractSmoothingGradient3DCartesianView(F_vectors_ interactions)
Key point angle ← ExtractKeyPointAngle(F_vectors_ interactions)
Human skeleton features← ExtractHumanSkeletonFeatures(F_vectors_ interactions)
Feature-vectors ← GetFeaturevector
Context-aware Feature-vectors.append (F_vectors)
End
Sense-aware Feature-vectors ← Normalize (sense-awarefeature_vectors)
return sense –aware feature vectors(sf_1,sf_2,sf_3, . . . ,sf_n)

(**a**) (**b**)

Figure 11. Moveable body parts: (**a**) detected human moveable body parts, and (**b**) binary view of detected human moveable body parts over SVW dataset.

3.5. Feature Optimization: Charged System Search Algorithm

For extracted features data optimization, we used a charged system search (CSS) [45] algorithm which is based on some defined principles of applied sciences. Charged system search (CSS) utilized the two laws from the applied sciences, namely, a Newtonian law from mechanics and Coulomb's law from physics, where Coulomb's law defines the electric force's magnitude between two charged points. Equation (17) defines the mathematical representation of Coulomb's law as

$$C_{ij} = n_e \frac{p_i \, p_j}{a^2{}_{ij}} \tag{17}$$

where C_{ij} denotes Coulomb's equation, a_{ij} represents the distance between two charged points and n_e is Coulomb's constant. Suppose a solid insulating sphere with a radius of r

and holding a true positive charge p_i, and f_{ij} as the outer side of the insulating sphere is considered as an electric field, which is defined as:

$$f_{ij} = n_e \frac{p_i}{a^2{}_{ij}} \tag{18}$$

CSS utilizes the concept of charged particles (CP); every CP creates an electric field using its magnitude property, which is denoted as p_i. The magnitude of a CP is defined as:

$$p_i = \frac{fit(i) - fitt_{worst}}{fitt_{best} - fitt_{worst}}, \; i = 1, 2 \ldots\ldots\ldots, N, \tag{19}$$

where $fit(i)$ is defined as the objective function of CSS, N is the limits of CP, while $fitt_{worst}$ and $fitt_{best}$ are for worst fitness declaration for all participants and $fitt_{best}$ for the best so far. The distance between two charged points is defined as:

$$a_{ij} = \frac{\| M_i - M_j \|}{\| M_i - M_j)/2 - M_{best} \| + \mathcal{E}} \tag{20}$$

where both M_i and M_j are the ith and jth location of the CPs, respectively; M_{best} denotes the CP's best position; and $\mathcal{E}$ is used to avoid uniqueness. Figure 12 shows the flowchart for the charged system search (CSS). Figure 13 represents a few results over three different classes of the different datasets.

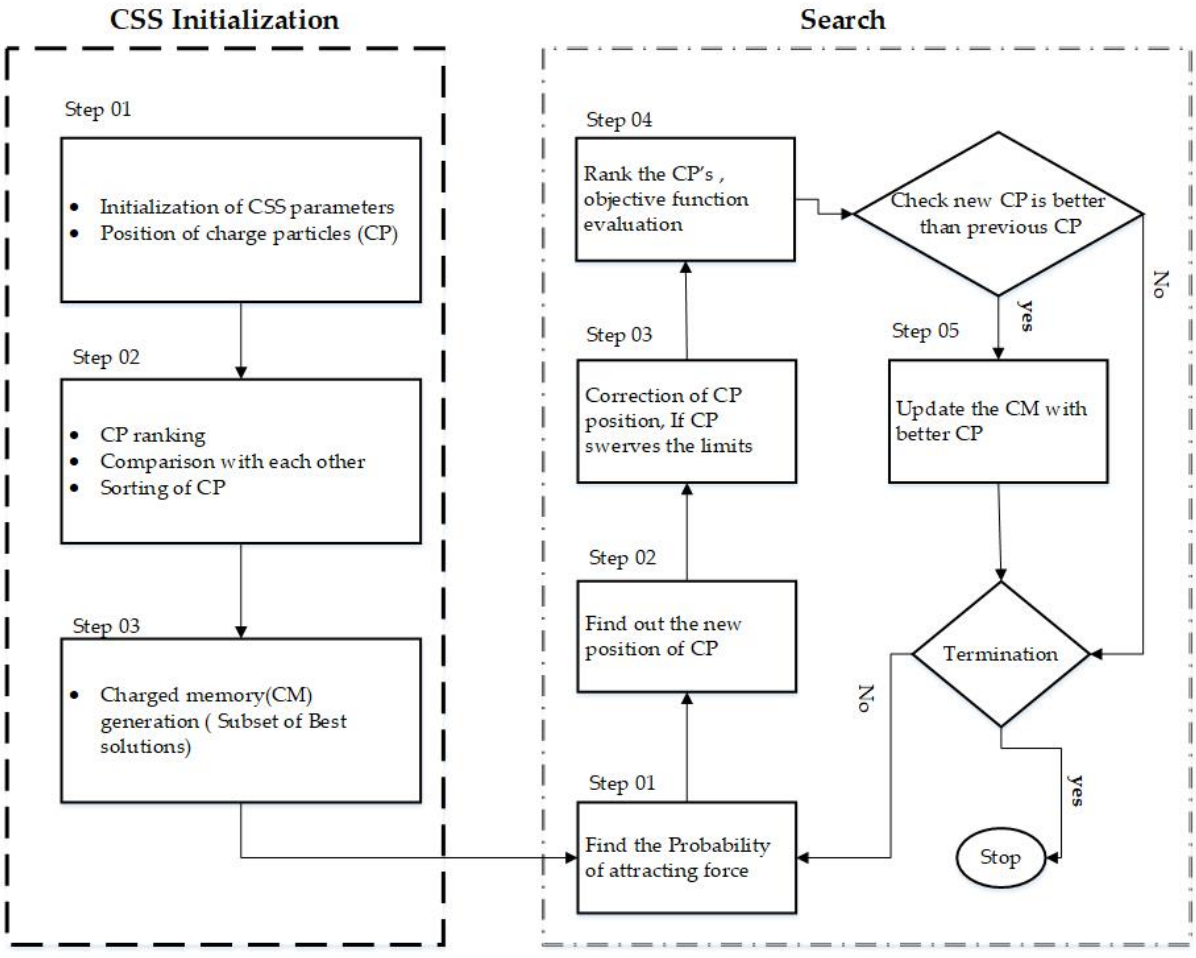


Figure 12. Charged system search (CSS) algorithm flow chart.

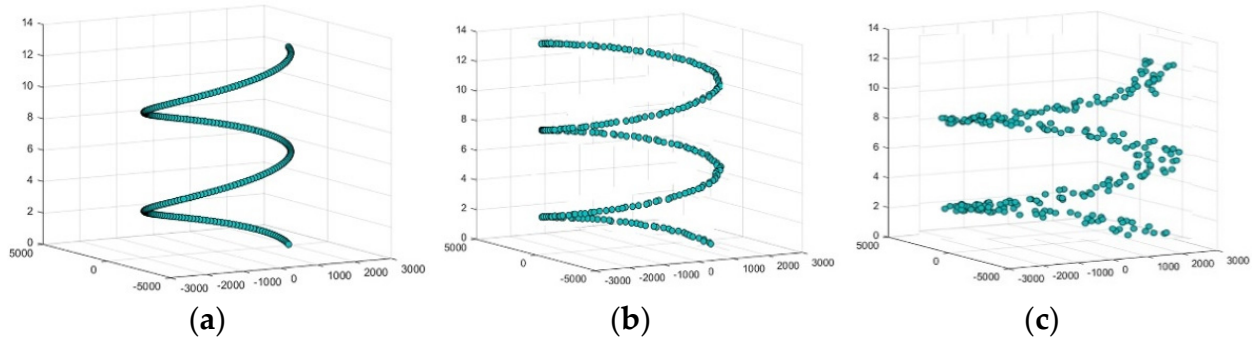


Figure 13. A few examples of charged system search (CSS) optimization algorithm results: (**a**) results over the kicking class of the UT-interaction dataset; (**b**) results over the Olympic sports dataset; (**c**) results over the football class of the SVW dataset.

3.6. Event Classification Engine: Deep Belief Network

In this section, we describe the machine learning-based deep belief network (DBN) [46], which we used as an event classifier. We used DBN over three datasets: SVW, Olympic Sports and UT-interaction. For the construction of DBN, the general building block is Restricted Boltzmann Machine (RBN). A hidden and visible unit of layer RBN constitutes a two-layer structure. The combined energy configuration of both units is defined as

$$\begin{aligned} Enr(VI, HI, \theta) =& -\sum_{i=1}^{D} bV_i v_i - \sum_{j=1}^{F} a_{Hj} h_j - \sum_{i=1}^{D}\sum_{j=1}^{F} w_j VI_i HI_j \\ =& > -b^T VI - a^T HI - VI^T WHI \end{aligned} \tag{21}$$

where $\theta = \{bV_i, aH_j, we_{ij}\}$; we_{ij} denotes the weight among visible component i and hidden component j; bV_i and aH_j present the bias condition of the hidden and visible components, respectively. The combined unit's configuration is defined as

$$Pr(VI, HI, \theta) = \frac{1}{NC(\theta)} exp(-Enr(VI, HI, \theta)) \tag{22}$$

$$NC(\theta) = \sum_{VI}\sum_{HI} Enr(VI, HI, \theta) \tag{23}$$

where $NC(\theta)$ denotes a regularization constant. The energy function is used as a probability distribution to the network; using Equation (21), the training vector can be adjusted. To extract the features from the data, the individual hidden layer of RBN is not a wise approach. The output of the first layer is used as the input of the second layer, and the output of the second layer is the input of the third layer of RBN. This hierarchal layer-by-layer structure of RBN develops the DBN; the deep feature extraction from the input dataset is more effective using a hierarchal approach of DBN. Figure 14 represents the graphical model and the general overview of DBN.

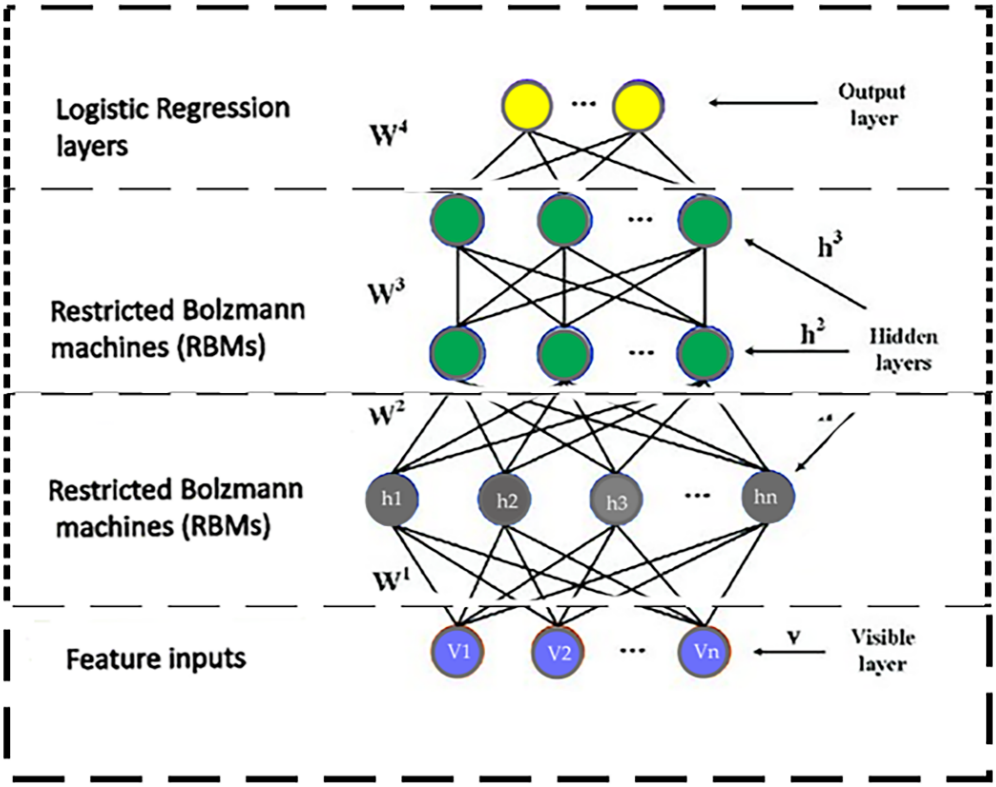


Figure 14. Layer model of the deep belief network.

4. Experimental Results

In this section, initially, we describe three different publicly available challenging datasets. After the description of the three datasets, we represent three types of tentative results. Exploration of the human body point recognition accuracies with distances to their ground truth was considered in the first experiment. After that, the next experiment was based on stochastic remote sensing event classification accuracies. Finally, in the last experiment, we compared event classification accuracies as well as human body part recognition accuracies with other well-known statistical state-of-the-art systems.

4.1. Datasets Description

The Olympic sports dataset [51] images for bowling, discus throw, diving_platform_10m, hammer throw, javelin throw, long jump, pole vault, shot put, snatch, basketball lay-up, triple jump and vault are event-based classes shot at a size of 720 × 480, 30 fps throughout the video. Figure 15 shows some example images of the Olympic sports dataset.

The UT-interaction dataset includes videos of six classes of continuously executed human–human encounters: shake-hands, point, embrace, drive, kick and strike. A sample of 20 video streams with a duration of about 1 min was available. The increasing video data involve at least another execution every encounter, giving us an average of eight iterations of human interactions per video. Numerous respondents participate throughout the video clips with even more than 15 distinct wardrobe situations. The images were shot at a size of 720 × 480, 30 fps throughout the video. There are six different interaction classes: handshaking, hugging, kicking pointing, punching and pushing. Figure 16 gives some example images from the UT-interaction dataset.

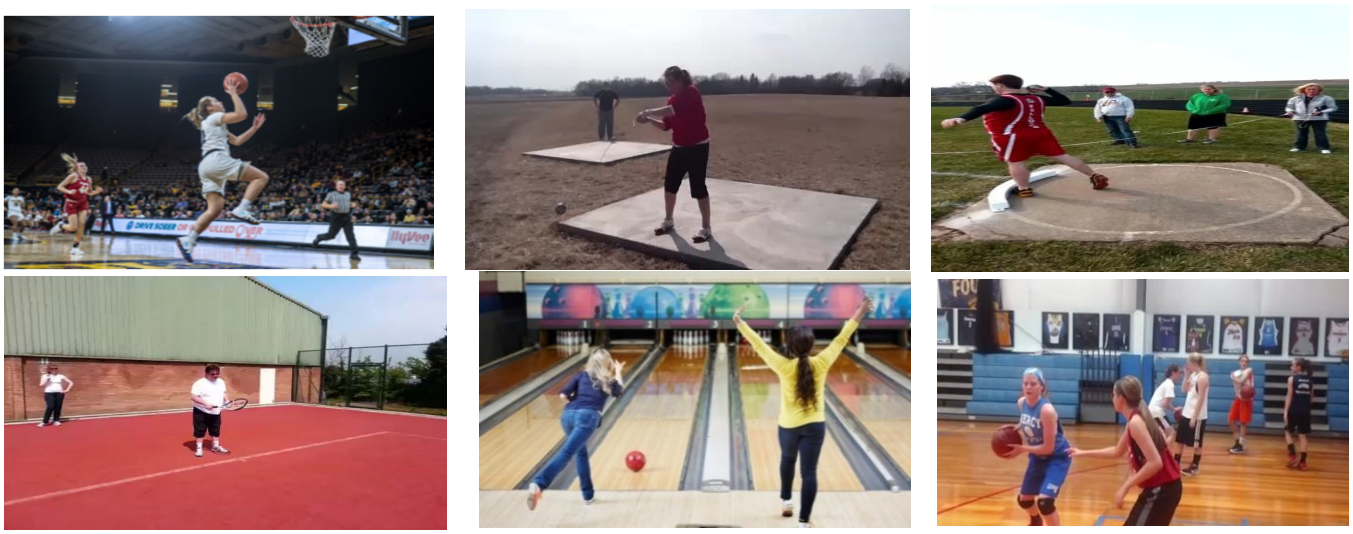


Figure 15. Some examples of multiview Olympic sports dataset.

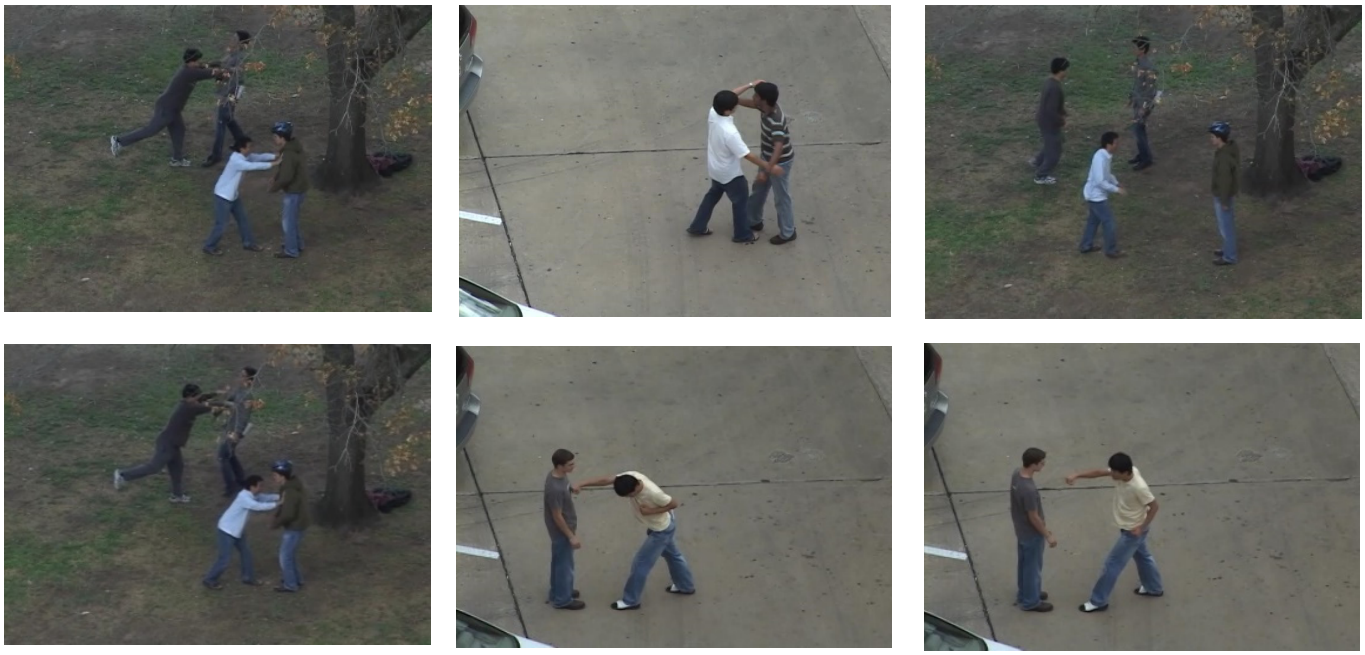


Figure 16. Some examples of UT-interaction dataset.

Sports Videos in the Wild (SVW) [52] 4200 were shot using the Coach's Eye mobile app, a pioneering sport development app produced by the TechSmith organization exclusively for the smartphone. There are nineteen event-based classes of 19 different events, namely, archery, baseball, basketball, BMX, bowling, boxing, cheerleading, football, golf, high jump, hockey, hurdling, javelin, long jump, pole vault, rowing, shotput, skating, tennis, volleyball and weight-lifting; the images were shot at a size of 720 × 480, 30 fps throughout. Figure 17 shows some example images from the SVW dataset.

Figure 17. A few samples of the SVW dataset.

The UCF aerial action dataset is based on a remote sensing technique. For video collection they used mini-drones from 400–450 feet. Five to six people are in the UCF aerial action dataset. They perform different events, such as walking, running and moving. Figure 18 shows some example images from the UCF aerial action dataset.

Figure 18. A few samples of the UCF aerial action dataset.

4.2. Experiment I: Body Part Detection Accuracies

To compute the efficiency and accuracy of human body part recognition, we estimated the distance [53,54] from the ground truth (GT) of the datasets using the following equation.

$$Di = \sqrt{\sum_{n=1}^{N} \left(\frac{I_n}{S_n} - \frac{J_n}{S_n} \right)^2} \tag{24}$$

Here, J is the GT of datasets and I is the position of the identified human body part. The threshold of 15 was set to recognize the accuracy between the identified human body part information and the GT data. Using the following Equation (25), the ratio of the identified human body parts enclosed within the threshold value of the categorized dataset was identified as

$$DA = \frac{100}{n} \left[\sum_{n=1}^{K} \begin{cases} 1 & if\ D \leq 15 \\ 0\ if\ D > 15 \end{cases} \right] \tag{25}$$

In Table 2, columns 2, 4 and 6 present the distances from the dataset ground truth and columns 3, 5 and 7 show the human body part recognition accuracies over the UT-interaction, Olympic sports and SVW datasets, respectively.

Table 2. Human body key point detection accuracy.

Body Key Points	Distance	UT (%)	Distance	Olympic Sports (%)	Distance	SVW (%)
HP	10.2	90	10	87	9	89
NP	9.8	85	11	83	10.1	86
RSP	10.5	80	11.1	81	12.1	83
REP	10.1	82	10.7	81	12.9	84
RWP	8.3	77	10.5	80	13	79
RHP	11.8	83	12	79	11	81
LSP	12.1	81	12.9	78	12	83
LEP	11	79	11	80	10	75
LWP	12.1	76	10	81	13.1	80
LHP	10.4	81	13	80	12.8	79
MP	11.1	90	14	89	12.9	87
RHP	12.9	77	13.9	80	11.1	83
LHP	11.1	80	10.1	75	10.8	80
LKP	11.2	93	11.8	90	9.3	94
RKP	10.9	90	12.3	87	12.7	89
RAP	11.5	81	12.7	80	12.6	79
LAP	11.2	78	13.1	79	13	78
LFP	12.3	95	11	94	11.1	92
RFP	10.5	90	9.3	93	10.8	91
Mean Accuracy Rate		83.57		83.00		83.78

HP = Head point, NP = Neck point, RSP = Right shoulder point, REP = Right elbow point, RWP = Right wrist point, RHP = Right hand point, LSP = Lift shoulder point, LEP = Left elbow point, LWP = Left wrist point, LHP = Left hand point, MP = Mid-point, RHP = Right hip point, LHP = Left hip point, LKP = left knee point, RKP = Right knee point, RAP = Right ankle point, LAP = Left ankle point, LFP = Left foot point, RFP = Right foot point.

Table 3 shows the key body points results of multiperson tracking accuracy for the UCF aerial action dataset. For detected parts, we used ✓, and for failures we used ✗. We achieved accuracy for person1—73.1%, person2—73.6%, person3—073.7%, person4—73.8%, person5—63.1%, and a mean accuracy of 71.41%.

Table 3. Human body key points results of multiperson tracking accuracy over UCF aerial action dataset.

Body Parts	Person1	Person2	Person3	Person4	Person5
HP	✓	✓	✓	✓	✓
NP	✓	✓	✓	✓	✓
RSP	✓	✗	✓	✗	✗
REP	✗	✓	✓	✓	✓
RWP	✓	✗	✓	✗	✓
RHP	✓	✓	✗	✓	✓
LSP	✗	✓	✗	✓	✗
LEP	✓	✗	✓	✓	✗
LWP	✓	✓	✓	✓	✓
LHP	✗	✗	✓	✗	✗
MP	✓	✓	✓	✓	✓
RHP	✓	✓	✗	✗	✓
LHP	✓	✓	✓	✓	✗
LKP	✗	✓	✓	✗	✓
RKP	✓	✗	✗	✓	✗
RAP	✓	✓	✓	✓	✓

Table 3. *Cont.*

Body Parts	Person1	Person2	Person3	Person4	Person5
LAP	✗	✓	✗	✓	✗
LFP	✓	✓	✓	✓	✓
RFP	✓	✓	✓	✓	✓
Accuracy	73.1%	73.6%	73.7%	73.8%	63.1%
Mean accuracy = 71.41%					

HP = Head point, NP = Neck point, RSP = Right shoulder point, REP = Right elbow point, RWP = Right wrist point, RHP = Right hand point, LSP = Lift shoulder point, LEP = Left elbow point, LWP = Left wrist point, LHP = Left hand point, MP = Mid-point, RHP = Right hip point, LHP = Left hip point, LKP = left knee point, RKP = Right knee point, RAP = Right ankle point, LAP = Left ankle point, LFP = Left foot point, RFP = Right foot point.

Table 4 shows the multiperson tracking accuracy over UCF aerial action dataset, column 1 shows the number of sequences, and each sequence has 25 frames. Column 2 shows the actual people of the dataset, Column 3 shows the successfully detected by over proposed system, column 4 shows the failure and finally, column 5 shows the accuracy and the mean accuracy is 91.15%.

Table 4. Multiperson tracking accuracy for the UCF aerial action dataset.

Sequence No (Frames = 25)	Actual Track	Successful	Failure	Accuracy
6	4	4	0	100
12	4	4	0	100
18	5	5	0	100
24	6	5	1	87.33
30	6	5	1	83.43
Mean Accuracy = 94.15				

4.3. Experiment II: Event Classification Accuracies

For stochastic remote sensing event classification, we used a deep belief network as an event classifier, and the proposed system was evaluated by the Leave One Subject Out (LOSO) cross-validation technique. In Figure 19, the results over the UT-interaction dataset show 91.67% event classification accuracy.

After this, we applied the deep belief network over the Olympic sports dataset and found the stochastic remote sensing event classification results. Figure 20 shows the results of the confusion matrix of event classification over the Olympic sports dataset with 92.50% mean accuracy.

Finally, we applied a deep belief network over the SVW dataset, and we found 89.47% mean accuracy for event classification. Figure 21 shows the confusion matrix of the SVW dataset with 89.47% mean accuracy.

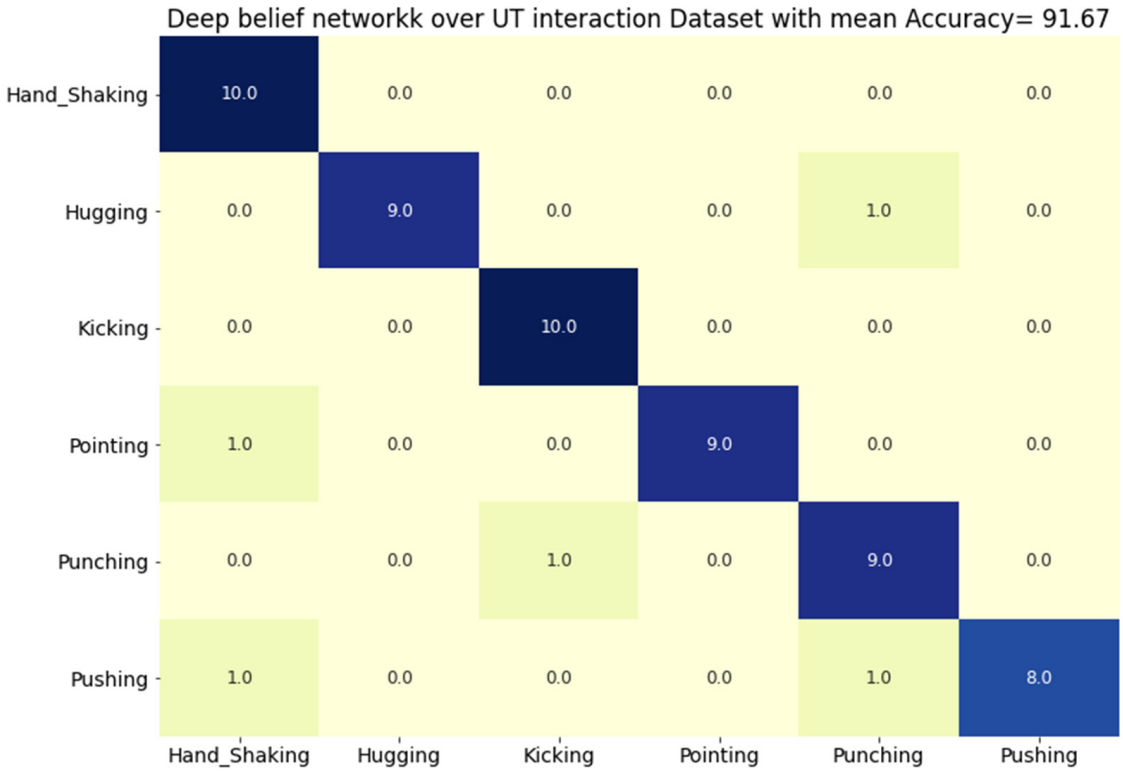


Figure 19. Confusion matrix of the proposed method for UT-interaction dataset.

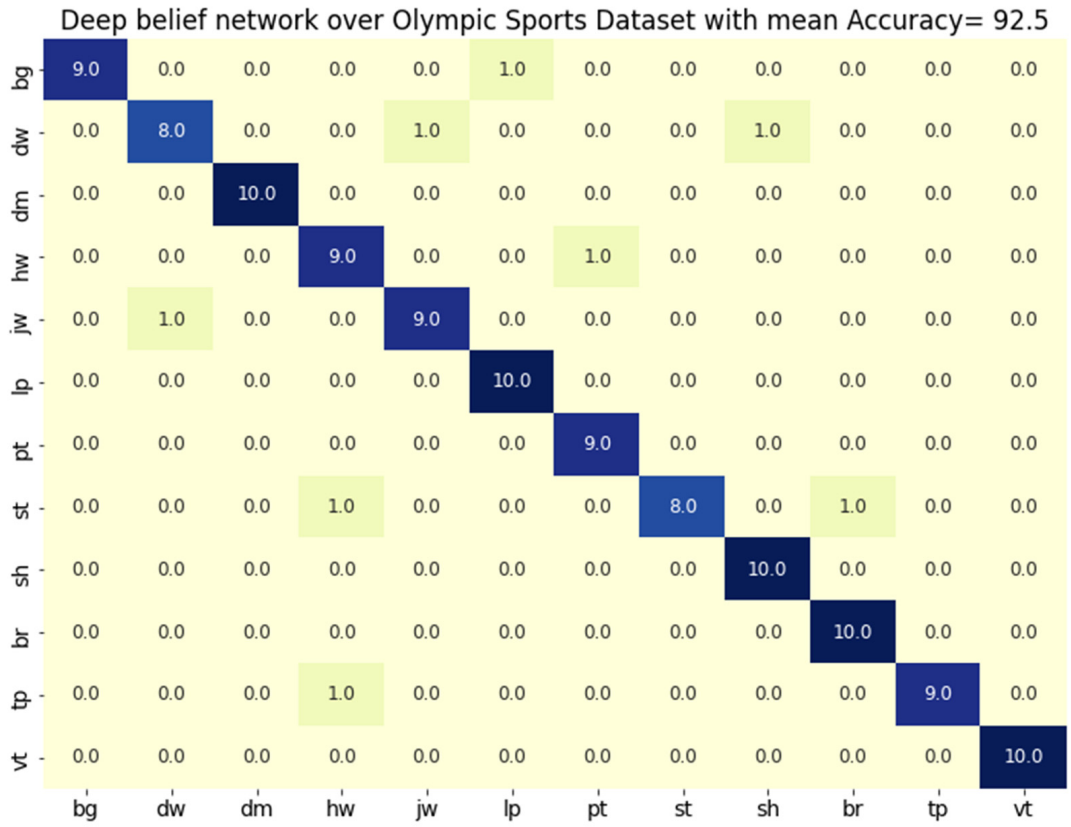


Figure 20. Confusion matrix of the proposed method for the Olympic sports dataset. bg = bowling, dw = discus throw, dm = diving_platform_10m, hw = hammer throw, jw = javelin throw, lp = long jump, pt = pole vault, st = shot put, sh = snatch, br = basketball lay-up, tp = triple jump, vt = vault.

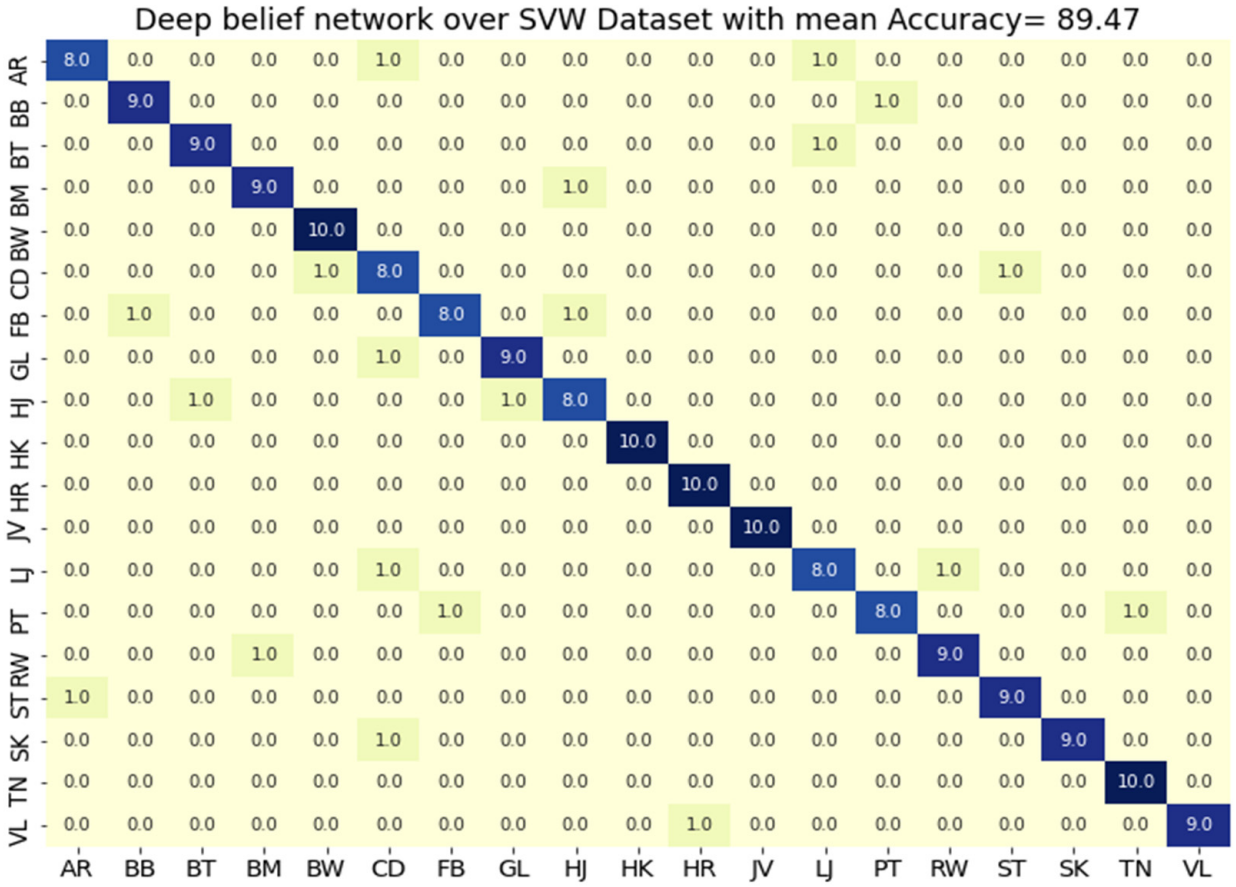


Figure 21. Confusion matrix of the proposed method for the SVW dataset. AR = archery, BB = baseball, BT = basketball, BM = bmx, BW = bowling, CD = cheerleading, FB = football, GL = golf, HJ = highjump, HK = hockey, HR = hurdling, JV = javelin, LJ = longjump, PT = polevault, RW = rowing, ST = shotput, SK = skating, TN = tennis, VL = volleyball.

4.4. Experiment III: Comparison with Other Classification Algorithms

In this section, we compare the precision, recall and f-1 measure over the SVW dataset, Olympic sports dataset and UT-interaction dataset. For the classification of stochastic remote sensing events, we used an Artificial Neural Network and Adaboost, and we compared the results with the deep belief network. Table 5 shows the results over the UT-interaction dataset, Table 6 shows the results over the Olympic sports dataset and Table 7 shows the results over the SVW dataset.

Table 5. Well-known classifiers comparison by considering precision, recall and F-1 measure over UT-interaction dataset.

Event Classes	Artificial Neural Network			Adaboost			Deep Belief Network		
Events	Precision	Recall	F-1 Measure	Precision	Recall	F-1 Measure	Precision	Recall	F-1 Measure
HS	0.727	0.800	0.762	0.778	0.875	0.824	0.833	1.000	0.909
HG	0.875	0.875	0.875	1.000	0.875	0.933	1.000	0.900	0.947
KI	0.889	0.727	0.800	0.889	0.889	0.889	1.000	1.000	1.000
PT	0.700	0.875	0.778	0.875	0.875	0.875	0.900	0.900	0.900
PN	0.700	0.778	0.737	0.875	0.778	0.824	0.900	0.900	0.900
PS	0.727	0.727	0.727	0.875	0.778	0.824	0.889	0.800	0.842
Mean	0.770	0.797	0.780	0.882	0.845	0.861	0.920	0.917	0.916

HS = hand shaking, HG = hugging, KI = kicking, PT = pointing, PN = punching, PS = pushing.

Table 6. Well-known classifiers comparison by considering precision, recall and F-1 measure over Olympic sports dataset.

Event Classes	Artificial Neural Network			Adaboost			Deep Belief Network		
Events	Precision	Recall	F-1 Measure	Precision	Recall	F-1 Measure	Precision	Recall	F-1 Measure
BG	0.538	0.636	0.583	0.600	0.750	0.667	1.000	0.900	0.947
DW	0.692	0.750	0.720	0.636	0.636	0.636	0.889	0.800	0.842
DM	0.700	0.583	0.636	0.750	0.750	0.750	1.000	1.000	1.000
HW	0.636	0.636	0.636	0.636	0.636	0.636	0.900	0.900	0.900
JW	0.727	0.889	0.800	0.889	0.889	0.889	0.900	0.900	0.900
LP	0.778	0.636	0.700	0.800	0.667	0.727	0.909	1.000	0.952
PT	0.615	0.800	0.696	0.778	0.778	0.778	0.900	1.000	0.947
ST	0.563	0.750	0.643	0.563	0.750	0.643	1.000	0.800	0.889
SH	0.875	0.700	0.778	1.000	0.778	0.875	0.909	1.000	0.952
BR	0.583	0.700	0.636	0.778	0.778	0.778	0.909	1.000	0.952
TP	0.750	0.750	0.750	0.700	0.700	0.700	1.000	1.000	1.000
VT	0.750	0.667	0.706	0.733	0.636	0.681	1.000	1.000	1.000
Mean	0.684	0.708	0.690	0.739	0.729	0.730	0.943	0.942	0.940

BG = bowling, DW = discus throw, DM = diving_platform_10m, HW = hammer throw, JW = javelin throw, LP = long jump, PT = pole vault, ST = shot put, SH = snatch, BR = basketball lay-up, TP = triple jump, VT = vault.

Table 7. Well-known classifiers comparison by considering precision, recall and F-1 measure over SVW dataset.

Event Classes	Artificial Neural Network			Adaboost			Deep Belief Network		
Events	Precision	Recall	F-1 Measure	Precision	Recall	F-1 Measure	Precision	Recall	F-1 Measure
AR	0.636	0.778	0.700	0.778	0.778	0.778	0.889	0.800	0.842
BB	0.667	0.727	0.696	0.700	0.700	0.700	0.900	0.900	0.900
Bt	0.667	0.727	0.696	0.615	0.727	0.667	0.900	0.900	0.900
BM	0.727	0.727	0.727	0.800	0.615	0.696	0.900	0.900	0.900
BW	0.692	0.750	0.720	0.750	0.750	0.750	0.909	1.000	0.952
CD	0.538	0.583	0.560	0.533	0.727	0.615	0.667	0.800	0.727
FB	0.700	0.700	0.700	0.750	0.692	0.720	0.889	0.800	0.842
GL	0.800	0.615	0.696	0.700	0.778	0.737	0.900	0.900	0.900
HJ	0.636	0.700	0.667	0.583	0.636	0.609	0.800	0.800	0.800
HK	0.800	0.889	0.842	0.889	0.889	0.889	1.000	1.000	1.000
HR	0.889	0.533	0.667	0.800	0.800	0.800	0.909	1.000	0.952
JV	0.800	0.727	0.762	1.000	0.727	0.842	1.000	1.000	1.000
LI	0.636	0.778	0.700	0.538	0.636	0.583	0.800	0.800	0.800
PT	0.875	0.700	0.778	0.875	0.700	0.778	0.889	0.800	0.842
RW	0.727	0.889	0.800	0.778	0.778	0.778	0.900	0.900	0.900
ST	0.778	0.700	0.737	0.778	0.700	0.737	0.900	0.900	0.900
SK	0.500	0.500	0.500	1.000	0.778	0.875	1.000	0.900	0.947
TN	0.615	0.615	0.615	0.600	0.818	0.692	0.909	1.000	0.952
VL	0.778	0.875	0.824	0.875	0.700	0.778	1.000	0.900	0.947
Mean	0.709	0.711	0.704	0.755	0.733	0.738	0.851	0.853	0.851

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

4.5. Experimentation IV: Qualitative Analysis and Comparison of our Proposed System with State-of-the-Art Techniques

Table 8 represents the qualitative analysis and comparison with existing state-of-the-art methods. Columns 1 and 2 show the comparison of human body part detection; columns 3 and 4 show the comparison results of human posture estimation; columns 5 and 6 represent the comparisons for stochastic remote sensing event classification. Results show a significant improvement in the proposed method.

Table 8. Qualitative analysis and comparison of the state-of-the-art methods with the proposed APEEC method.

Proposed by	Body Part Detection Accuracy (%)	Proposed by	Human Posture Estimation Mean Accuracy (%)	Proposed by	Event Classification Mean Accuracy (%)
S. Hong, et al. [55]	76.60	Chen et al. [56]	95.0	Mahmood et al. [57]	83.50
C. Dorin et al. [58]	81.90	Zhang et al. [59]	96.0	Amer, M.R. et al. [60]	82.40
S. Gomathi et al. [61]	82.71	Li et al. [62]	98.0	Zhu, Y [63]	83.10
Proposed	83.78	Proposed	98.3	Proposed	92.50

4.6. Experimentation V: Comparison of our Proposed System with State-of-the-Art Techniques

In this section, we compare the proposed system with existing state-of-the-art methods, and we check the mean accuracy of stochastic remote sensing event classification and human body part detection. Table 6 shows the comparison results with existing state-of-the-art methods. The results show the superior performance of our proposed Adaptive Posture Estimation and Event Classification (APEEC) system. Because nineteen body parts are considered, a pseudo-2D stick model, multifused data, data optimization via CSS and event classification are evaluated using DBN. In [64], Rodriguez et al. present a novel technique via the most sensible human movement in which they used vibrant descriptions and tailored loss mechanisms to inspire a reproductive framework to find precise future human movement estimates. Xing et al. [65] designed a fusion feature extraction framework, in which they syndicate mutually stationary features and dynamic features to cover additional action material from video data. In [66], Chiranjoy et al. developed a supervised method for automatic identification the, key contribution being the extraction of spatiotemporal features, and they spread the vectors of locally aggregated descriptors (VLADs) as a dense video encoding demonstration. In [67], S. Sun et al. proposed an approach for feature extraction in which they extract directed optical flow along with a CNN-based model for human event identification and classification. In [68], Reza. F et al. defined an approach to deal with event identification and classification with the CNN and Network in Network architecture (NNA), which are the baseline of modern CNN. The lightweight architecture of CNN, average, max and product functions are used to identify human events. In [69], L. Zhang et al. [49] designed an innovative framework for human-based video event identification and classification via binary-level neural network learning. At the initialization stage, CNN is used to recognize the main video content. Finally, they extract spatiotemporal features via Gated Recurrent Unit (GRU) and Long Short Term Memory (LSTM)-based methods. Wang. H et al. [70] developed a human movement approximation approach in which they improve dense features using a video-based camera. For the multifused data, they consider optical flow and Speed-up Robust features (SURF). A. Nadeem et al. [71] designed a novel framework for human posture estimation via a multidimensional feature vector, human body point identification. For recognition, they used the Markov entropy model, while Quadratic discriminant analysis (QDA) was used for feature extraction from video data. Mahmood et al. [57] developed a novel human activity, event and interaction detection model for human-based video data. They applied the segmentation process to extract the human silhouette and multistep human body parts, which are based on points to base distance features to recognize the events. In [60], Amer, M.R. et al. proposed a unified approach for human activity recognition using spatiotemporal-based data features. In [63], Kong, Y. et al. designed a robust human event-based technique in which they used human local and global body part multidata features to recognize human-based events and interactions. Table 9 shows a comprehensive comparison of our proposed Adaptive Posture Estimation and Event Classification (APEEC) method with state-of-the-art methods:

Table 9. Comparison of the state-of-the-art methods with the proposed APEEC method.

Methods	UT-Interaction (%)	Methods	Olympic Sports (%)	Methods	Sports Videos in the Wild (%)
C. Rodriguez et al. [64]	71.80	Zhang.L et al. [69]	59.10	S. Sun et al. [67]	74.20
Xing et al. [65]	85.67	Sun. S et al. [67]	74.20	Reza. F et al. [68]	82.30
Chiranjoy. Cet et al. [66]	89.25	Wang. H et al. [70]	89.60	A. Nadeem et al. [71]	89.09
Wang. H et al. [70]	89.10	A. Nadeem et al. [71]	88.26	Zhu, Y [72]	83.10
Mahmood et al. [57]	83.50	E. Park et al. [73]	89.10	—	—
Amer, M.R. et al. [60]	82.40	M. Jain et a [74]	83.20	—	—
Kong, Y. et al. [63]	85.00	—	—	—	—
Proposed method	91.67		92.50		89.47

5. Discussion

The proposed APEEC was designed to achieve Stochastic Remote Sensing Event Classification over adaptive human posture estimation. In this approach, we extracted multifused data from human-based video data; after that, layered data optimization via a charged system search algorithm and event classification using a deep belief network were conducted. The proposed method starts with input video data; for that, we used three publicly available datasets in video format. A preprocessing step was performed to reduce noise. First, we used adaptive filters, which have high computational complexity. To reduce computational cost, we applied a Gaussian filter to reduce noise. Video to frame conversion and resizing of the extracted frames also help to save time and memory. The next step is human detection, which was performed by GMM and a Saliency map algorithm. After successfully extracting human silhouettes, we found the human body key points that are located on the upper and lower body. This is the baseline for adaptive human posture estimation, which is based on the unbreakable pseudo-2D stick model.

The next step is multifused data; we extracted two types of features: first, full human body features: conations energy, moveable body parts and 3D Cartesian view features, and second, key point features: skeleton zigzag and geometric features. To overcome resource costs, we adopted a data optimization technique in which we used the Met heuristic data optimization charged system search algorithm. Finally, we applied a deep belief network for stochastic remote sensing event classification.

We faced some limitations and problems in the APEEC system. We were not able to find the hidden information for the human silhouette, and this is the reason for the low accuracy of human posture analysis and stochastic remote sensing event classification. Figure 22 shows some examples of problematic events. In the images, we can see the skeleton and human body key point locations; however, the positions are not clear due to complex data and occlusion of some points of the human body.

Here, we present results and analysis of the proposed APEEC. The mean accuracy for human body part detection is 83.57% for the UT-interaction dataset, 83.00% for the Olympic sports dataset and 83.78% for the SVW dataset. Mean event classification accuracy is 91.67% over the UT-interaction dataset, 92.50% for the Olympic sports dataset and 89.47% for SVW dataset. These results are superior in comparison with existing state-of-the-art methods.

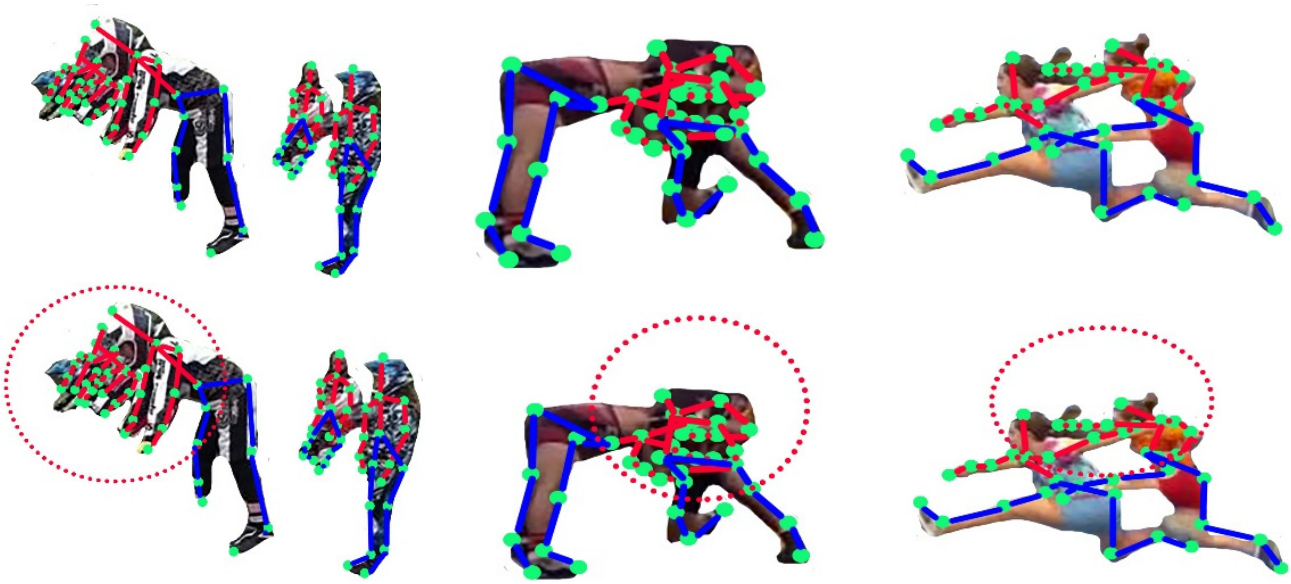


Figure 22. Examples of problematic results over the SVW dataset.

6. Conclusions

We contribute a robust method for the detection of nineteen human body parts during complex human movement, challenging events and human postures, which can be detected and estimated more accurately than other methods. To achieve more accurate results in adaptive posture estimation and classification, we designed a skeletal pseudo-2D stick model that enables the detection of nineteen human body parts. In the multifused data, we extracted sense-aware features, which include energy, moveable body parts, skeleton zigzag features, angular geometric features and 3D Cartesian features. Using these extracted features, we can classify events into multiple human-based videos more accurately. For data optimization, a hierarchical optimization model was implemented to reduce computational cost and to optimize data. Charged system search optimization was implemented with over-extracted features. A deep belief network was applied for multiple human-based video event classification.

6.1. Theoretical Implications

The proposed APEEC system works in different and complex scenarios to classify stochastic remote sensing events. APEEC works with multihuman-based datasets as well, although there are theoretical implications to determining the more complex application of the system in terms of event detection in videos, sports, medical, emergency services, hospital management system and surveillance system; however, for these applications, we can apply the proposed APEEC system in a real-time video data-capturing environment.

6.2. Research Limitations

The proposed APEEC system, the Sports Video in the Wild dataset, is a more complex dataset compared to the Olympic sports dataset and the UT-interaction dataset. Due to complex angle information and complex human information, we faced minor differences in results. Figure 22 presents the results of human posture detection, while the dotted circle highlights the occlusion and overlapping issues in a certain area. We faced difficulties when dealing with these types of data and environments. In the future, we will work on this problem by using a deep learning approach, and we will devise a new method to obtain outstanding results.

Author Contributions: Conceptualization, I.A. and K.K.; methodology, I.A. and A.J.; software, I.A.; validation, M.G.; formal analysis, M.G. and K.K.; resources, M.G., A.J. and K.K.; writing—review and editing, I.A., A.J. and K.K.; funding acquisition, A.J. and K.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (No. 2018R1D1A1A02085645).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tahir, S.B.U.D.; Jalal, A.; Kim, K. Wearable Inertial Sensors for Daily Activity Analysis Based on Adam Optimization and the Maximum Entropy Markov Model. *Entropy* **2020**, *22*, 579. [CrossRef] [PubMed]
2. Tahir, S.B.U.D.; Jalal, A.; Batool, M. Wearable Sensors for Activity Analysis using SMO-based Random Forest over Smart home and Sports Datasets. In Proceedings of the 2020 3rd International Conference on Advancements in Computational Sciences (ICACS), Lahore, Pakistan, 17–19 February 2020; 2020; pp. 1–6.
3. Susan, S.; Agrawal, P.; Mittal, M.; Bansal, S. New shape descriptor in the context of edge continuity. *CAAI Trans. Intell. Technol.* **2019**, *4*, 101–109. [CrossRef]
4. Rehman, M.A.U.; Raza, H.; Akhter, I. Security Enhancement of Hill Cipher by Using Non-Square Matrix Approach. In Proceedings of the 4th international conference on knowledge and innovation in Engineering, Science and Technology, Berlin, Germany, 21–23 December 2018.
5. Tingting, Y.; Junqian, W.; Lintai, W.; Yong, X. Three-stage network for age estimation. *CAAI Trans. Intell. Technol.* **2019**, *4*, 122–126. [CrossRef]
6. Wiens, T. Engine speed reduction for hydraulic machinery using predictive algorithms. *Int. J. Hydromech.* **2019**, *2*, 16. [CrossRef]
7. Shokri, M.; Tavakoli, K. A review on the artificial neural network approach to analysis and prediction of seismic damage in infrastructure. *Int. J. Hydromech.* **2019**, *2*, 178. [CrossRef]
8. Jalal, A.; Sarif, N.; Kim, J.T.; Kim, T.-S. Human Activity Recognition via Recognized Body Parts of Human Depth Silhouettes for Residents Monitoring Services at Smart Home. *Indoor Built Environ.* **2012**, *22*, 271–279. [CrossRef]
9. Jalal, A.; Uddin, Z.; Kim, T.-S. Depth video-based human activity recognition system using translation and scaling invariant features for life logging at smart home. *IEEE Trans. Consum. Electron.* **2012**, *58*, 863–871. [CrossRef]
10. Jalal, A.; Kim, Y.; Kim, D. Ridge body parts features for human pose estimation and recognition from RGB-D video data. In Proceedings of the Fifth International Conference on Computing, Communications and Networking Technologies (ICCCNT), Hefei, China, 11–14 July 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 1–6.
11. Jalal, A.; Kamal, S.; Kim, D. A Depth Video Sensor-Based Life-Logging Human Activity Recognition System for Elderly Care in Smart Indoor Environments. *Sensors* **2014**, *14*, 11735–11759. [CrossRef]
12. Akhter, I. Automated Posture Analysis of Gait Event Detection via a Hierarchical Optimization Algorithm and Pseudo 2D Stick-Model. Ph.D. Thesis, Air University, Islamabad, Pakistan, December 2020.
13. Jalal, A.; Nadeem, A.; Bobasu, S. Human Body Parts Estimation and Detection for Physical Sports Movements. In Proceedings of the 2019 2nd International Conference on Communication, Computing and Digital systems (C-CODE), Islamabad, Pakistan, 6–7 March 2019; pp. 104–109.
14. Mahmood, M.; Jalal, A.; Kim, K. WHITE STAG model: Wise human interaction tracking and estimation (WHITE) using spatio-temporal and angular-geometric (STAG) descriptors. *Multimed. Tools Appl.* **2020**, *79*, 6919–6950. [CrossRef]
15. Quaid, M.A.K.; Jalal, A. Wearable sensors based human behavioral pattern recognition using statistical features and reweighted genetic algorithm. *Multimed. Tools Appl.* **2020**, *79*, 6061–6083. [CrossRef]
16. Nadeem, A.; Jalal, A.; Kim, K. Human Actions Tracking and Recognition Based on Body Parts Detection via Artificial Neural Network. In Proceedings of the 3rd International Conference on Advancements in Computational Sciences (ICACS 2020), Lahore, Pakistan, 17–19 February 2020; pp. 1–6.
17. Ahmed, A.; Jalal, A.; Kim, K. A Novel Statistical Method for Scene Classification Based on Multi-Object Categorization and Logistic Regression. *Sensors* **2020**, *20*, 3871. [CrossRef] [PubMed]
18. Jalal, A.; Mahmood, M. Students' behavior mining in e-learning environment using cognitive processes with information technologies. *Educ. Inf. Technol.* **2019**, *24*, 2797–2821. [CrossRef]
19. Gochoo, M.; Tan, T.-H.; Huang, S.-C.; Batjargal, T.; Hsieh, J.-W.; Alnajjar, F.S.; Chen, Y.-F. Novel IoT-based privacy-preserving yoga posture recognition system using low-resolution infrared sensors and deep learning. *IEEE Internet Things J.* **2019**, *6*, 7192–7200. [CrossRef]
20. Gochoo, M.; Tan, T.-H.; Liu, S.-H.; Jean, F.-R.; Alnajjar, F.S.; Huang, S.-C. Unobtrusive Activity Recognition of Elderly People Living Alone Using Anonymous Binary Sensors and DCNN. *IEEE J. Biomed. Heal. Informatics* **2018**, *23*, 1. [CrossRef] [PubMed]
21. Lee, M.W.; Nevatia, R. Body Part Detection for Human Pose Estimation and Tracking. In Proceedings of the 2007 IEEE Workshop on Motion and Video Computing (WMVC'07), Austin, TX, USA, 23–24 February 2007; p. 23.
22. Aggarwal, J.; Cai, Q. Human Motion Analysis: A Review. *Comput. Vis. Image Underst.* **1999**, *73*, 428–440. [CrossRef]
23. Wang, L.; Hu, W.; Tan, T. Recent developments in human motion analysis. *Pattern Recognit.* **2003**, *36*, 585–601. [CrossRef]

24. Liu, J.; Luo, J.; Shah, M. Recognizing realistic actions from videos "in the Wild". In Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009, Miami Beach, FL, USA, 20–25 June 2009.
25. Khan, M.A.; Javed, K.; Khan, S.A.; Saba, T.; Habib, U.; Khan, J.A.; Abbasi, A.A. Human action recognition using fusion of multiview and deep features: An application to video surveillance. *Multimedia Tools Appl.* **2020**, 1–27. [CrossRef]
26. Zou, Y.; Shi, Y.; Shi, D.; Wang, Y.; Liang, Y.; Tian, Y. Adaptation-Oriented Feature Projection for One-shot Action Recognition. *IEEE Trans. Multimedia* **2020**, 1. [CrossRef]
27. Franco, A.; Magnani, A.; Maio, D. A multimodal approach for human activity recognition based on skeleton and RGB data. *Pattern Recognit. Lett.* **2020**, *131*, 293–299. [CrossRef]
28. Ullah, A.; Muhammad, K.; Haq, I.U.; Baik, S.W. Action recognition using optimized deep autoencoder and CNN for surveillance data streams of non-stationary environments. *Futur. Gener. Comput. Syst.* **2019**, *96*, 386–397. [CrossRef]
29. Jalal, A.; Kamal, S.; Kim, D.-S. Detecting Complex 3D Human Motions with Body Model Low-Rank Representation for Real-Time Smart Activity Monitoring System. *KSII Trans. Internet Inf. Syst.* **2018**, *12*, 1189–1204. [CrossRef]
30. Jalal, A.; Mahmood, M.; Hasan, A.S. Multi-features descriptors for Human Activity Tracking and Recognition in Indoor-Outdoor Environments. In Proceedings of the 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST), Islamabad, Pakistan, 8–12 January 2019; pp. 371–376.
31. van der Kruk, E.; Reijne, M.M. Accuracy of human motion capture systems for sport applications; state-of-the-art review. *Eur. J. Sport Sci.* **2018**, *18*, 806–819. [CrossRef] [PubMed]
32. Wang, Y.; Mori, G. Multiple Tree Models for Occlusion and Spatial Constraints in Human Pose Estimation. In *European Conference on Computer Vision*; the Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer International Publishing: Berlin/Heidelberg, Germany, 2008; Volume 5304, pp. 710–724.
33. Amft, O.O.; Tröster, G. Recognition of dietary activity events using on-body sensors. *Artif. Intell. Med.* **2008**, *42*, 121–136. [CrossRef] [PubMed]
34. Wang, Y.; Du, B.; Shen, Y.; Wu, K.; Zhao, G.; Sun, J.; Wen, H. EV-Gait: Event-Based Robust Gait Recognition Using Dynamic Vision Sensors. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 6351–6360.
35. Jiang, Y.-G.; Dai, Q.; Mei, T.; Rui, Y.; Chang, S.-F.; Chang, S.-F. Super Fast Event Recognition in Internet Videos. *IEEE Trans. Multimedia* **2015**, *17*, 1. [CrossRef]
36. Li, A.; Miao, Z.; Cen, Y.; Zhang, X.-P.; Zhang, L.; Chen, S. Abnormal event detection in surveillance videos based on low-rank and compact coefficient dictionary learning. *Pattern Recognit.* **2020**, *108*, 107355. [CrossRef]
37. Einfalt, M.; Dampeyrou, C.; Zecha, D.; Lienhart, R. Frame-Level Event Detection in Athletics Videos with Pose-Based Convolutional Sequence Networks. In *Proceedings of the 2nd International Workshop on Multimedia Content Analysis in Sports—MMSports'19*; Association for Computing Machinery (ACM): New York, NY, USA, 2019; pp. 42–50.
38. Yu, J.; Lei, A.; Hu, Y. Soccer Video Event Detection Based on Deep Learning. In *Proceedings of the Constructive Side-Channel Analysis and Secure Design*; Springer International Publishing: Berlin/Heidelberg, Germany, 2018; pp. 377–389.
39. Franklin, R.J.; Mohana; Dabbagol, V. Anomaly Detection in Videos for Video Surveillance Applications using Neural Networks. In Proceedings of the 2020 Fourth International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 8–10 January 2020; pp. 632–637.
40. Lohithashva, B.; Aradhya, V.M.; Guru, D. Violent Video Event Detection Based on Integrated LBP and GLCM Texture Features. *Rev. d'intelligence Artif.* **2020**, *34*, 179–187. [CrossRef]
41. Feng, Q.; Gao, C.; Wang, L.; Zhao, Y.; Song, T.; Li, Q. Spatio-temporal fall event detection in complex scenes using attention guided LSTM. *Pattern Recognit. Lett.* **2020**, *130*, 242–249. [CrossRef]
42. Khan, M.H.; Zöller, M.; Farid, M.S.; Grzegorzek, M. Marker-Based Movement Analysis of Human Body Parts in Therapeutic Procedure. *Sensors* **2020**, *20*, 3312. [CrossRef] [PubMed]
43. Esfahani, M.I.M.; Zobeiri, O.; Moshiri, B.; Narimani, R.; Mehravar, M.; Rashedi, E.; Parnianpour, M. Trunk Motion System (TMS) Using Printed Body Worn Sensor (BWS) via Data Fusion Approach. *Sensors* **2017**, *17*, 112. [CrossRef]
44. Golestani, N.; Moghaddam, M. Human activity recognition using magnetic induction-based motion signals and deep recurrent neural networks. *Nat. Commun.* **2020**, *11*, 1–11. [CrossRef]
45. Kaveh, A.; Talatahari, S. A novel heuristic optimization method: Charged system search. *Acta Mech.* **2010**, *213*, 267–289. [CrossRef]
46. Chen, Y.; Zhao, X.; Jia, X. Spectral–Spatial Classification of Hyperspectral Data Based on Deep Belief Network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* **2015**, *8*, 2381–2392. [CrossRef]
47. Jalal, A.; Kamal, S.; Kim, D. Depth silhouettes context: A new robust feature for human tracking and activity recognition based on embedded HMMs. In Proceedings of the 2015 12th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), Goyang City, Korea, 28–30 October 2015; pp. 294–299.

48. Li, H.; Lu, H.; Lin, Z.L.; Shen, X.; Price, B.L. Inner and Inter Label Propagation: Salient Object Detection in the Wild. *IEEE Trans. Image Process.* **2015**, *24*, 3176–3186. [CrossRef] [PubMed]
49. Moschini, D.; Fusiello, A. Tracking Human Motion with Multiple Cameras Using an Articulated Model. In *Computer Graphics Collaboration Techniques and Applications*; Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer International Publishing: Berlin/Heidelberg, Germany, 2009; Volume 5496, pp. 1–12.
50. Jalal, A.; Akhtar, I.; Kim, K. Human Posture Estimation and Sustainable Events Classification via Pseudo-2D Stick Model and K-ary Tree Hashing. *Sustainability* **2020**, *12*, 9814. [CrossRef]
51. Niebles, J.C.; Chen, C.-W.; Fei-Fei, L. Modeling Temporal Structure of Decomposable Motion Segments for Activity Classification. In *Proceedings of the Constructive Side-Channel Analysis and Secure Design*; Springer International Publishing: Berlin/Heidelberg, Germany, 2010; pp. 392–405.
52. Safdarnejad, S.M.; Liu, X.; Udpa, L.; Andrus, B.; Wood, J.; Craven, D. Sports Videos in the Wild (SVW): A video dataset for sports analysis. In Proceedings of the 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Ljubljana, Slovenia, 4–8 May 2015; Volume 1, pp. 1–7.
53. Wang, L.; Zhang, Y.; Feng, J. On the Euclidean distance of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1334–1339. [CrossRef] [PubMed]
54. Akhter, I.; Jalal, A.; Kim, K. Pose Estimation and Detection for Event Recognition using Sense-Aware Features and Ada-boost Classifier. In Proceedings of the 18th International Bhurban Conference on Applied Sciences and Technology (IBCAST), Islamabad, Pakistan, 12–16 January 2021.
55. Hong, S.; Kim, M. A Framework for Human Body Parts Detection in RGB-D Image. *J. Korea Multimedia Soc.* **2016**, *19*, 1927–1935. [CrossRef]
56. Chen, X.; Yuille, A. Articulated pose estimation by a graphical model with image dependent pairwise relations. *arXiv* **2014**, arXiv:1407.3399.
57. Mahmood, M.; Jalal, A.; Sidduqi, M.A. Robust Spatio-Temporal Features for Human Interaction Recognition Via Artificial Neural Network. In Proceedings of the 2018 International Conference on Frontiers of Information Technology (FIT), Islamabad, Pakistan, 17–19 December 2018; pp. 218–223.
58. Dorin, C.; Hurwitz, B. Automatic body part measurement of dressed humans using single rgb-d camera. In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; pp. 3042–3048.
59. Zhang, D.; Shah, M. Human pose estimation in videos. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 2012–2020.
60. Amer, M.R.; Todorovic, S. Sum Product Networks for Activity Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 800–813. [CrossRef] [PubMed]
61. Gomathi, S.; Santhanam, T. Application of Rectangular Feature for Detection of Parts of Human Body. *Adv. Comput. Sci. Technol.* **2018**, *11*, 43–55.
62. Li, Y.; Liu, S.G. Temporal-coherency-aware human pose estimation in video via pre-trained res-net and flow-CNN. In Proceedings of the International Conference on Computer Animation and Social Agents (CASA), Seoul, Korea, 22–24 May 2017; pp. 150–159.
63. Kong, Y.; Liang, W.; Dong, Z.; Jia, Y. Recognising human interaction from videos by a discriminative model. *IET Comput. Vis.* **2014**, *8*, 277–286. [CrossRef]
64. Rodriguez, C.; Fernando, B.; Li, H. Action Anticipation by Predicting Future Dynamic Images. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
65. Xing, D.; Wang, X.; Lu, H. Action Recognition Using Hybrid Feature Descriptor and VLAD Video Encoding. In *Asian Conference on Computer Vision*; Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer International Publishing: Berlin/Heidelberg, Germany, 2015; Volume 9008, pp. 99–112.
66. Chattopadhyay, C.; Das, S. Supervised framework for automatic recognition and retrieval of interaction: A framework for classification and retrieving videos with similar human interactions. *IET Comput. Vis.* **2016**, *10*, 220–227. [CrossRef]
67. Sun, S.; Kuang, Z.; Sheng, L.; Ouyang, W.; Zhang, W. Optical Flow Guided Feature: A Fast and Robust Motion Representation for Video Action Recognition. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1390–1399.
68. Rachmadi, R.F.; Uchimura, K.; Koutaki, G. Combined convolutional neural network for event recognition. In Proceedings of the Korea-Japan Joint Workshop on Frontiers of Computer Vision, Takayama, Japan, 17–19 February 2016; pp. 85–90.
69. Zhang, L.; Xiang, X. Video event classification based on two-stage neural network. *Multimedia Tools Appl.* **2020**, *79*, 21471–21486. [CrossRef]
70. Wang, H.; Oneata, D.; Verbeek, J.; Schmid, C. A Robust and Efficient Video Representation for Action Recognition. *Int. J. Comput. Vis.* **2016**, *119*, 219–238. [CrossRef]
71. Nadeem, A.; Jalal, A.; Kim, K. Accurate Physical Activity Recognition using Multidimensional Features and Markov Model for Smart Health Fitness. *Symmetry* **2020**, *12*, 1766. [CrossRef]
72. Zhu, Y.; Zhou, K.; Wang, M.; Zhao, Y.; Zhao, Z. A comprehensive solution for detecting events in complex surveillance videos. *Multimedia Tools Appl.* **2018**, *78*, 817–838. [CrossRef]

73. Park, E.; Han, X.; Berg, T.L.; Berg, A.C. Combining multiple sources of knowledge in deep CNNs for action recognition. In Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 7–10 March 2016.
74. Jain, M.; Jegou, H.; Bouthemy, P. Better Exploiting Motion for Better Action Recognition. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013.