Syntactic Model-Based Human Body 3D Reconstruction and Event Classification via Association based Features Mining and Deep Learning

Yazeed Yasin Ghadi[1], Israr Akhter[2], Mohammed Alarfaj[3], Ahmad Jalal[2], and Kibum Kim[4]

[1] Department of Computer Science and Software Engineering, Al Ain University, Al Ain 15551, UAE; Yazeed.ghadi@aau.ac.ae

[2] Department of Computer Science, Air University, 46000, Islamabad, Pakistan; israrakhter.edu@gmail.com, ahmadjalal@mail.au.edu.pk

[3] Department of Electrical Engineering, King Faisal University, Hofuf-31982, Saudi Arabia; mkalarfaj@kfu.edu.sa [4] Department of Human-Computer Interaction, Hanyang University, Ansan, 15588, South Korea

kibum@hanyang.ac.kr;

Corresponding Author:

Kibum Kim[4]

Department of Human-Computer Interaction, Hanyang University, Ansan, 15588, South Korea

kibum@hanyang.ac.kr;

ABSTRACT

The study of human posture analysis and gait event detection from various types of inputs is a key contribution to the human life log. With the help of this research and technologies humans can save costs in terms of time and utility resources. In this paper we present a robust approach to human posture analysis and gait event detection from complex video-based data. For this, initial posture information, landmark information are extracted, and human 2D skeleton mesh are extracted, using this information set we reconstruct the human 2D to 3D model. Contextual features, namely, degrees of freedom over detected body parts, joint angle information, periodic and non-periodic motion, and human motion direction flow, are extracted. For features mining, we applied the rule-based features mining technique and, for gait event detection and classification, the deep learning-based CNN technique is applied over the mpii-video pose, the COCO, and the pose track datasets. For the mpii-video pose dataset, we achieved a human landmark detection mean accuracy of 87.09% and a gait event recognition mean accuracy of 90.90%. For the COCO dataset, we achieved a human landmark detection mean

accuracy of 87.36 % and a gait event recognition mean accuracy of 89.09%. For the pose track dataset, we achieved a mean accuracy of 87.09% in human landmark detection and a gait event recognition mean accuracy of 88.18%. The proposed system performance shows a significant improvement compared to existing state-of-the-art frameworks.

Subjects Artificial Intelligence, Computer Vision, 2D/3D reconstruction, Deep learning, Machine learning

Keywords 2D to 3D Reconstruction, Convolutional Neural Network, Gait Event Classification, Human Posture Analysis, Landmark Detection, Synthetic Model, Silhouette Optimization.

INTRODUCTION

With regard to human posture information, motion estimation, and gait event detection from various types of input such as camera-based data, sensor-based datasets currently provide the most challenging issues. Various approaches and models are proposed to find more accurate and appropriate methods and functions for event classification and human body posture, motion and movement analysis. Data generation, communication, and transmission are routine takes for smart systems such as hospital management systems, educational systems, emergency systems, communication systems *ud din Tahir, Jalal & Kim; ud din Tahir (2020)*, store records, airport or other transportation systems *Jalal, Khalid & Kim, 2020; Khalid et al (2021).* The generated data is needed to be processed and utilized in the context of finding some useful information for humans in terms of time-saving, and reducing the cost of manpower *Jalal, Khalid & Kim (2020).* Human posture and gait event analysis can help us represent human movement as information that is useful in smart systems for the detection and identification of human events and conditions such as standing, walking, running, playing, singing and dancing.

Various smart systems such as smart surveillance systems, cryptography in smart systems, and data security systems *ur Rehman, Raza & Akhter (2018)*, management systems and smart sports systems save us time and manpower and thus also money. With smart systems, many aspects of the behavior and condition of patients can be monitored automatically and more efficiently, thus relieving the burden on medical staff and patient care and freeing them from some redundancies. For sports, we can recognize the current event such as classification about games. Smart security systems can detect the security status, security issues, normal and abnormal events in places such as airports, or defense facilities.

Following are the main key points of problem statements.

- In various crowd-based systems, human motion analysis is a challenge to perform and archive.
- For human motion analysis, mostly 2D human model was suggested and, in this regard, 3D reconstruction is proposed.

- For data optimization, most of the research was based on traditional data optimization techniques. To deal with this, data mining techniques are proposed.
- Event recognition is challenging in the computer vision domain, especially when complex human crowd-based data is used as an input of the system. To tackle this problem, we adopted contextual features and CNN-based classification.

Technical Gaps:

In this domain, the various model was proposed with various techniques. Many papers deal with human motion analysis and gait event detection separately. In the above papers, research gaps and technical gaps are related throughout input-output. Initially, input is based on video-based data, and pre-pressing is performed to deal with data processing cast and time-saving. After that, the motion analysis of human detection, human landmark, human 2D stick model, and 3D human reconstruction are performed. This technique is robust to find and analyze the human motion information data which was normally dealt with in previous papers and was based on the gap of human motion methods. For event recognition, the pre-step to find the features with a data mining approach using CNN gait event detection and recognition is performed. On the other hand, previous papers deal with various models and data optimization which was the issue of processing cast and time.

Contribution of this Paper:

Thus, in this research article, we propose a robust method for human body posture analysis and human gait event detection. For this, we use the mpii-video-pose dataset; the COCO dataset and the pose track dataset as input, and initially preprocess the video samples for frame conversion, motion blur noise reduction and resizing. The next step in human detection is to process these data sources by various algorithms for silhouette extraction, optimization of detected human silhouettes and human body landmark detection. Then, the next stage is to analyze the human posture information. For this, given 2D images to approximate 3D image reconstruction is applied using ellipsoid and synthetic modeling of the human body. This is followed by the feature extraction phase in which contextual features information is extracted. This information includes the degree of freedom, periodic motion, non-periodic motion, motion direction, flow, and rotational and angular joint features. For features mining, an association-based technique is adopted. For gait event classification, CNN is applied.
The main contribution of this paper is:

- With complex datasets, the detection of and optimization of humans and the extraction of and optimization of human silhouettes are challenging. We, therefore, propose a robust human silhouette extraction approach.
- For human motion, posture, and movement information analysis, we propose a method for the conversion of the human skeleton-based 2D mesh to the 3D human skeleton.

- For the detection and recognition of gait events and contextual features, the extraction approaches are proposed in which degrees of freedom (DOF), periodic motion, non-periodic motion, motion direction, flow, and rotational angular joint features are extracted.
- Finally, data mining and classification via hierarchical methods, mining, and CNN-based methods are adopted for gait event classification.

The subdivisions of this article are as follows: we start with related works, followed by our system methodology, then, the detailed experimental setup discussion and, finally, an overview of the paper is presented in the conclusion.

RELATED WORK

Innovations in smartphone cameras and recorded video and developments in object marker sensor-based devices allow for more efficient farming and collection of data for exploration and research in the area. Several novel and effective approaches for recognizing human events, movements, and postures have been developed in the past. Table 1 includes a comprehensive review of recent research in this area.

MATERIALS & METHODS

For the video input of our proposed method, the main source is RGB cameras which provide clean and noise-free data. The first step is video to frame conversion 30 fps which reduces computational cost and time. After this, noise reduction techniques are applied. Human detection is achieved using Markov random field, change detection, floor detection, and spatial-temporal dereferencing. Then, 3D human reconstruction is achieved in which computational models with ellipsoids, the synthetic model with supper quadrics, joint angle estimation, and 3D reconstruction are applied. After this contextual features extraction is applied. For features mining we applied association-based techniques. Finally, gait event classification is achieved with the help of CNN over three state-of-the-art datasets. Figure 1 demonstrates the proposed system model's structural design.

Figure 1. **The proposed system model's structural design.**

The complete overview of the proposed method is described in (Algorithm 1) in which steps of the proposed method are described and main functions with sub-functions and equations are mentions.

The graphical superstations show the flow of every program, system, or software. Figure 2 shows the graphical repetition in the shape of a flow chart of the proposed methodology.

Figure 2. Flow chart of the proposed method.

173
174

Preprocessing of the data

175
176   Before the detection of human body landmarks, some preprocessing
177 methods are applied to save computational cost and time. Initially, video data
178 is converted into images and then a motion blur filter is applied to reduce
179 excess information.

Background subtraction

180
181   For background subtraction, we applied an optimized merging method
182 technique in which we initially applied Markov random field based on color
183 information and region merging methods. After this, change detection in
184 image sequence is applied over an adaptive threshold-based approach, floor
185 detection, and finally spatial-temporal differencing is adopted to get more
186 accurate results. Figure 2 shows the results of background subtraction
187 techniques.

188
189
190

Figure 3. Results of different background subtraction techniques along with the original image. (a) Original image (b) change detection (c) floor detection (d) Markov random field and (e) spatial-temporal differencing.

Silhouette optimization and human detection

191
192   In this sub-phase of landmark detection, we find the optimized human
193 silhouette through the merging of change detection, floor detection, Markov
194 random field, and spatial-temporal differencing techniques with the help of
195 an adaptive threshold approach. (Algorithm 2) shows the detailed procedure
196 of silhouette optimization.

197
198
199   After this, human detection is performed in two phases, initially, head
200 detection is performed with the help of a human head "size and shape-based"
201 technique. We set the weight of a human head as w0=1/25 of the human
202 silhouette and, using region of interest model, we find the super pixel position
203 of the human body and, after this, Gaussian kernel is used to capture the likely
204 area of the human head. Finally, using this human head information, human
205 shape and appearance information, human body movement, and motion
206 information, human detection and identification is performed. Eq. (1) is used
207 for head tracking

208
209
$$T_{He}^{q} \leftarrow T_{He}^{q-1} + \Delta T_{He}^{q-1} \tag{1}$$

210
211 where T_{He}^{q} represents a human head land-mark location in any given video
212 frame q which is consequential to calculating by the frame differences. For
213 human detection, Eq. (2) shows the mathematical relationship.

214

$$T_{FH}^{q} = (T_{He}^{q} \leftarrow T_{He}^{q-1} + \Delta T_{He}^{q-1}) + T_{End}^{q} \tag{2}$$

where T_{FH}^{q} represents a human location in any given video frame q and T_{End}^{q} shows the bounding box size for human detection. Figure 3 shows the results of optimized human body silhouettes, head detection, and human detection.

(a)   (b)   (c)

Figure 4. Results of (a) Optimized human silhouette (b) human head detection (c) human detection in RGB videos and image sequences.

Once human silhouette extraction and human detection are achieved, the next phase is to find human body landmarks for the posture estimation and analysis of the human body movements.

Body landmarks detection

In this sub-phase of landmark detection, we establish human body landmarks using a fast marching algorithm; we have applied this to the full human silhouette. Initially, the center point of the human body is extracted for the distance value $dis(h) = 0$, where h is the initial point and is distinguished as a marked point. All remaining unmarked points of the human body are considered as $dis(p) = \infty$. this process is applied to every detected point and to the pixel value of the human silhouette. The mathematical representation is:

$$dis = \left\{ \frac{dis_x + dis_y + \sqrt{\Delta}}{2} \right\} \text{ when } \Delta \geq 0 \tag{3}$$

$$\min(dis_x + dis_y) + wi \text{ otherwise} \tag{4}$$

$$\Delta = x^2 - (dis_x - dis_y)^2$$

where dis_x and dis_y is the geodesic distance in the 2d plane, correspondingly, $dis_y = min(Disi + 1, mo, Disi - 1, mo)$ and $dis_y = min(Disi, no + 1, Di, no - 1)$. After this, human body parts estimation is performed by finding the midpoint of the human body and the hands, elbows, neck, head, knees, and feet points are extracted *Gochoo et al. (2021)*. The detection of the human midpoint is represented as;

$$T_{t}^{q} \leftarrow T_{to}^{q-1} + \Delta T_{to}^{q-1} \tag{5}$$

where T_{t}^{q} represents a human midpoint location in any given video frame q which is consequential to calculation according to frame differences *Akhter, Jalal & Kim, (2021)*. To find the knees points we take the midpoint between the human-body midpoint and the two feet points. Eq. (6) demonstrates the human knee points;

$$T_{k}^{q} = \left(T_{t}^{q} - T_{f}^{q} \right) / 2 \tag{6}$$

where T_{k}^{q} is a knee point, T_{m}^{q} is the human body midpoint, and T_{f}^{q} denotes a foot point. For each elbow position estimation we utilized the neck point and respective (left/right) hand point information and found the mid point between the hand and the neck points. Eq. (7) as;

258

$$T_e^q = (T_{hn}^q - T_{nq}^q)/2 \qquad (7)$$

259   where T_e^q denotes the human elbow point, T_{hn}^q is the human hand point, and
260   T_{nq}^q denotes the neck point. Figure 5 represents the results of landmarks and
261   body parts.

262

263   (a)   (b)

Figure 5. Human body landmark detection results (a) presents the landmark results using an HSV color map, (b) presents the eleven human body points.

264   In this section, basic 2D human body skeletonization is achieved, using
265 extracted human body key points. Figure 6 represents the detailed view of the
266 human body 2D skeleton *Jalal, Akhtar & Kim (2020)* over eleven human
267 body parts. The human body 2D skeleton is based on three main body areas:
268 Upper body parts of the human body *Ubph*, Mid parts of the human body
269 *Mph*, and lower body parts of the human body *Lbph*. *Ubph* denote the
270 connection of the human head (h), neck point (nq), elbow (e), and human
271 hand points (hn). *Mph* is initiated through the human midpoint (t) and *Unph*.
272 *Lbph* is based on human knee points (hk) and footpoints (f). Every human
273 body part or joint takes a specific letter k to complete a particular action. Eq.
274 (8), (9) and (10) demonstrate the calculated associations of the human 2D
275 stick model as;

$$Ubph = h \bowtie nq \bowtie S\_e \bowtie S\_hn \qquad (8)$$
$$Mph = S\_t \bowtie Ubph \qquad (9)$$
$$HLbs = hk \bowtie f \bowtie Mph \qquad (10)$$

279
280

Figure 6. The Human 2D skeleton model results in over eleven human body parts.

281

3D human reconstruction

283   For the reconstruction of the 3D human shape, we utilized a human 2D
284 skeleton as a base step and as information for human posture estimation. For
285 the reconstruction of the 3D human shape, we used the human body parts
286 location and information.

287

A. Computational model with ellipsoids

289   In this subpart of the 3D human reconstruction, the first step is achieved
290 with a computational model with ellipsoid techniques. We take the joints
291 information, the head point is connected with the neck point by an ellipsoid
292 shape, and the neck point is connected with elbow points via an ellipsoid
293 shape. The Elbow point is connected with the hand via an ellipsoid shape and
294 the same procedure is followed for the remaining human body points.

$$C_{me} = P_a(e_x, e_y) \blacksquare P_{a+1}(e_x, e_y) \qquad (11)$$

296

where C_{me} is the computational model with ellipsoids, $P_a(e_x, e_y e_{z,})$ is the first point with the value of x,y and $P_{a+1}(e_x, e_y)$ is the next point of the human body with the value of x,y. Figure 7 shows the results of the computational model with ellipsoids over human body points.

Figure 7. The results of the computational model with ellipsoids over human body points.

After completion of the computational model with ellipsoids, the next phase is the synthetic model with super quadrics.

B. Synthetic model with super quadrics

To display the human posture and estimation of human motion the synthetic model with super quadrics is adopted, using the computational model with ellipsoids information we utilize the previous ellipsoids and convert them into a rectangular shape for more accurate information and analysis of human posture.

$$S_{SQ} = C_{me} \rightarrow s_e \tag{12}$$

Where S_{SQ} is the synthetic model with super quadrics and C_{me} is the computational model with ellipsoid information, $\rightarrow s_e$ is the reshaping of the given information of points. Figure 8 shows the results of the synthetic model with super quadrics over the computational model with ellipsoids.

(a)   (b)   (c)

Figure 8. The results of the synthetic model with super quadrics over human body points. (a) Human 2D skeleton, (b) computational model with ellipsoids (c) synthetic model with super quadrics.

C. Joint angle estimation

For 3D reconstruction of the ellipsoids, the prerequisite step is to estimate joint angle information. For this, volumetric data and the edge information of the human body key points are extracted. For movement and angle information using the global and local coordinate system, we estimate the DOF for human body key points root information. After that, the Cartesian product of the skeleton graph is estimated for further processing.

$$T_{Ne}^q \leftarrow T_{Ne}^{q-1} + \Delta T_{Ne}^{q-1} \tag{13}$$

where T_{Ne}^q represents the neck landmark location in any given video frame, q is consequential to calculation of the frame differences. See Figure. 9.

Figure 9. The theme concept of local and global coordinate systems. The left side shows the local coordinate system over the human left knee; the right side shows the DOF based global coordinate system.

$$F_v = \left[\theta_{g\_l}, \theta_{S\_h}, \theta_{S\_n}, \theta_{R\_e}, \theta_{L\_e}, \theta_{R\_h}, \theta_{L\_h}, \theta_{S\_m}, \theta_{R\_k}, \theta_{L\_k}, \theta_{R\_f}, \theta_{L\_f}\right] \quad (14)$$

where F_v represents the angle joint function, θ_{g_l} denotes the global to local coordinates, θ_{S_h} indicates the head point, θ_{S_n} shows the neck point, θ_{R_e} denotes the right elbow point, θ_{L_e} indicates the left elbow point, θ_{R_h} represents the right-hand point, θ_{L_h} shows the left-hand point, θ_{S_m} indicates the mid-point, θ_{R_k} shows the right knee point, θ_{L_k} denotes the left knee point, θ_{R_f} indicates the right foot point, θ_{L_f} shows the left foot point.

D. 3D ellipsoid reconstructions

Finally, the 3D reconstruction of the human body is implemented using human body joint information, ellipsoid information, and skeleton graphs. The preview of the 3D image gives us more precise and accurate posture information and estimation for further processing.

$$R_E(x,0 \mid I,D) \propto R_E(x)R_E(I|x)R_E(D|x)R_E(D|\theta) \quad (15)$$

Where $R_E(x,0 \mid I,D)$ is the 3D reconstruction ellipsoid, $\propto$ is the reshaping and $R_E(x)R_E(I|x)R_E(D|x)R_E(D|\theta)$ shows the angle information based on previous data. Figure 10 shows the results of the 3D ellipsoid reconstruction over the synthetic model with super quadrics and joint angle estimation.

Figure 10. The results of the 3D ellipsoid reconstruction over the synthetic model with super quadrics and joint angle estimation.

Contextual features extraction

In this section, the extraction of contextual features is implemented in which DOF, periodic motion, non-periodic motion, motion direction and flow and rotational angular joint features are extracted.

A. Degree of freedom

In contextual features extraction, Degree of freedom (DOF) is implemented over all body parts and the x,y,z dimension information In DOF features vector three directional angle values for each body parts, for knee points x_knee, y_knee, z_knee, for head points x_head, y_head, z_head, for neck points x_ neck, y_ neck, z_ neck, for elbow points x_ elbow, y_ elbow, z_ elbow, for hand points x_ hand, y_ hand, z_ hand, for midpoints x_ mid, y_ mid, z_ mid, for foot points x_ foot, y_ foot, z_ foot. Eq. (16) shows the mathematical relation for DOF.

368
$$D_{of} = A(\theta x, \theta y, \theta z) \uparrow \varepsilon D \qquad (16)$$

369 where D_{of} represents the degree of freedom feature vector, $\theta x, \theta y, \theta z$ shows
370 the three dimension of the angle and εD is the local and global coordinate
371 system. Figure 11 shows the results for the degree of freedom:

372

373

Figure 11. Few DOF results examples.

374

375 *B. Periodic motion*

376 In this contextual feature, human motion is detected over human body parts.
377 The targeted area of interest is the human body portion which provides
378 periodic motion. The detection of this area is performed with the base analysis
379 of the human body. A bounding box indicates the region of interest. The Eq.
380 (17) shows the mathematical relation for periodic motion.

381
$$PM(t) = \alpha \sin(\omega t + k) \qquad (17)$$

382 where $PM(t)$ denotes periodic motion and $\alpha \sin(\omega t + k)$ shows the relation
383 of human motion that is repeated in any given sequence of images. Figure 12
384 shows the results of periodic motion:

385

Figure 12. Periodic motion results.

386

387 *C. Non-periodic motion*

388 In the non-periodic motion contextual feature, human motion is detected over
389 human body parts. The targeted area of interest is the human body portion
390 which provides nonperiodic motion and non-uniform motion. The detection
391 of this area is performed with the base analysis of human body motion. A
392 bounding box indicates the region of interest. Eq. (18) shows the
393 mathematical relation of non-periodic motion:

394
$$NPm(t) = \| P_{t,t+1} - P_{t,t+2} \| \qquad (18)$$

395 where $NPm(t)$ denotes non-periodic motion and $P_{t,t+1} - P_{t,t+2}$ shows
396 the difference between the first and the next sequence of images in input data.
397 Figure 13 shows the results of non-periodic motion:

398

399

Figure 13. Results of non-periodic motion.

400

401

402 *D. Motion direction flow*

403 For the identification of more accurate gait events motion direction flow is
404 one a contributions in terms of contextual features. Using changes in motion
405 and human motion body flow we detect the direction of human body

movement and motion flow. Eq. (19) shows the mathematical model for motion direction flow features.

$$M_{df} = \Sigma_0^p I_{vl}(I) \rightarrow D \qquad (19)$$

where M_{df} is motion direction flow of the human body, I is the index values of the given image, I_{vl} is RGB (x,y,z) pixel indexes, and $\rightarrow D$ shows the motion direction. Figure 14 shows the results of motion direction flow over the basketball class video.

Figure 14. The results of motion direction flow over the basketball video.

E. Rotational angular joint

Rotational angular joint features are based on the angular geometry of human body parts. A 5 X 5 pixel region is used over detected body parts and from every node of the window of pixel region $\cos \theta$ is estimated and maps all values in the feature vector. The Eq. (20) shows the mathematical model of rotational angular joint features.

$$A1 = \cos (x,y) \rightarrow L, \; A2 = \cos (x,y) \rightarrow L \qquad (20)$$
$$A3 = \cos (x,y) \rightarrow L, \; A4 = \cos (x,y) \rightarrow L$$

where $A1, A2, A3, A4$ denotes the sides of the 5x5 windows, $\cos (x,y)$ represents the angle value over pixel x and y, and $\rightarrow L$ indicates the side to follow. Figure 15 shows the results of rotational angular joint features over the dance class.

Figure 15. Rotational angular joint results and the pattern of rotational angels.

After the completion of the contextual features portion, we concatenate all the sub-feature vectors into the main feature vector while (Algorithm 3) shows the detailed overview of the contextual features extraction approach.

Data optimization and features mining

The association rule-based features mining method helps us to pick the most unique features that screen out unnecessary and inconsistent features from the extracted dataset which tend to reduce gait event classification precision and accuracy. This is a bottom-up strategy that starts with a null feature set nf and progressively adds innovative features based on optimization function selection. This can decrease the mean square error which results in more significant details. The association rule-based features mining approach is

commonly used in various domains such as security systems, medical systems and image processing-based smart systems.

This technique helps to minimize the main features space data while the features mining approach is dependent upon the specific objective function for an optimal solution which plays the key role in gait event classification. The Bhattacharyya distance calculation features optimization approach is used in the present architecture for various human event-based classes. It can determine the differentiation rating $nf_{(x,y)}$ among different segments x and b and then test it.

$$nf_{(x,y)} = (g_x - g_y)(\frac{\Sigma_x - \Sigma_y}{2})(g_x - g_y)^t \qquad (21)$$

where $nf_{(x,y)}$ is the optimal features set, g_x are the mean and Σ_x are the covariance of class x and g_y are the mean and Σ_y are the covariance of class y for M numbers of event-based classes. The optimal solution score is computed as.

$$OFv = \frac{1}{N^2}\Sigma_{u=1}^{M}\Sigma_{v=1}^{M} nf_{(u,v)} \qquad (22)$$

A recognition assessment criterion is suggested for estimating different gait event classifications for an input dataset to acquire selected features that can be expected to eliminate classification errors as well as provide improved inter-class interpretability throughout features data. For the mpii-video-pose dataset DOF, Periodic motion, Rotational angular joint features are selected. For the Pose_track dataset, DOF, Motion direction and flow nonPeriodic motion features are selected. For the COCO dataset, Motion direction and flow nonPeriodic motion, Rotational angular joints features are selected. Figure 16 shows the most accurate features results over the mpii-video pose, the COCO, and the Pose track datasets.

Figure 16. The most accurate features results via the rule-based features mining approach over the mpii-video pose, COCO, and Pose track datasets.

Event classification

The extracted optimal features vector is used as input for Convolution Neural Network (CNN) for gait event classification. CNN is a deep learning-based classification approach which is widely used in image and video types of input data. CNN works well and it gives more accurate results compared to other traditional techniques. CNN adds less processing weight with minimum bias, thus providing a high accuracy rate.

The input, output, and hidden layers are the three main layers in a Convolutional Neural Network (CNN). The convolutional layer, important mechanisms, complete linked layer, and standardization layer are the four sub-divisions of each secret layer. The sub-band extracted features are across in the input neurons and highly correlated throughout the convolutional layer by a $5x5$ graded selector. The batch normalization method is then used to aggregate the responses among all neuronal populations. To further minimize feature dimensions, clustered solutions are convolutional and combined again, and then the interaction maps are computed by the completely connected sheet.

491
$$TR_p = \sum_q wi_{p.q} \times a_p + b_q \qquad (23)$$

492   where TR_p is the CNN transfer function, $wi_{p.q}$ is the connecting layer's
493   adjacent weight, a_p shows the input optimized features vectors and b_q is the
494   bias values. Furthermore, the regression algorithm is adopted for parameter
495   optimization and it reduces backpropagation errors. The output of the
496   regression algorithm $\sigma(TR_p)$ returns the distribution function of total
497   probability for possible n repetition over the output layer of CNN.

498
$$\sigma(TR_p) = \frac{e^{TR_p}}{\sum_{n=1}^{n} e^{TR_p}}, q = 1,..........n \qquad (24)$$

499   Figure 17 shows the detailed process of CNN parameters learning over gait
500   event detection and classification:

501

502

Figure 17. CNN model overview.

503

504

505   **RESULTS**

506   Dataset Descriptions

507   The Mpii-video pose data set is a large-scale dataset, which contains human
508   activities and posture information-based videos. 21 different activities such
509   as home activities, lawn, garden, sports, washing windows, picking fruit, and
510   rock climbing. All the videos All the videos selected for our dataset collection
511   have been recommended as YouTube top 10 videos in each activity. Figure
512   18 shows some example images of the Mpii-video-pose dataset.

513

Figure 18. A few example images of the mpii-video-pose dataset.

514

515   The COCO (Common Objects in Context) dataset is based on multi-person
516   tracking and object detection dataset, different activities contains in the
517   COCO dataset They include Bicycling, Conditioning exercise, Dancing,
518   Fishing and hunting, Music playing, Religious activities, Sports,
519   Transportation, Walking, Water activities, winter activities. Figure 19 shows
520   some example images of the COCO dataset:

521

522

Figure 19. A fe**w example images from the COCO dataset.**

523

524   Pose track dataset is based on two main tasks including multi-person human
525   pose estimation and analysis over a single frame. Videos and articulated
526   tracking have been based on human posture estimation. Dataset mostly
527   consists of complex videos such as crowded or team sports. Various activities

528   have been covered in the pose track dataset. Figure 20 shows some example
529   images of the pose track dataset.

530

531

Figure 20. A few example images of the postmark dataset.

532

533   Experiment I: The Landmarks Detection Accuracies

534   To calculate the effectiveness and precision of the detected body parts, we
535   approximate the geodesic distance *Akhter (2020) Akhter, Jalal & Kim*
536   *(2021)* from the given ground truth (*GT*) of the input datasets by Eq. (25):

537
$$Die = \sqrt{\sum_{n=1}^{M}\left(\frac{O_n}{P_n} - \frac{O_n}{P_n}\right)^2} \tag{25}$$

538   Here, O is the ground truth values of the datasets and P is the current
539   location of the recognized body part. The error margin of 16 is set to identify
540   the accuracy among the acknowledged body part value and the input data.
541   Through Eq. (26), the proportion of the recognized body parts encircled
542   within the error margin value of the considered data is known as;

543
$$De = \frac{100}{n}\left[\sum_{n=1}^{i}\left\{\begin{matrix} 1 & if & De \leq 16 \\ 0 & & if\ De > 16 \end{matrix}\right.\right] \tag{26}$$

544   In (Table 2), columns 2, 4, and 6 show the error distances from the given
545   dataset ground truth and columns 3, 5, and 7 show the body part recognition
546   and detection accuracies over the MPII, COCO and Posetrack datasets
547   respectively.

548

549   HP= Head point, NP= Neck point, REP= Right elbow point, RHP= Right
550   hand point, LEP= Left elbow point, LHP= Left hand point, MP= Mid-point,
551   LKP= left knee point, RKP= Right knee point, LFP= Left foot point,
552   RFP=Right foot point.
553   (Table 3) represents the results of multi-person human body parts for the mpii-
554   video-pose dataset. For identified body parts, we indicate with ✓ and for
555   unidentified we adopted we use✗. We attained a detection accuracy for
556   human1 -63.63%, human2 - 72.72%, human3 – 63.63%, human4- 72.72%,
557   human5 - 72.72% and the mean detection accuracy of 69.09%.

558

559   HP= Head point, NP= Neck point, REP= Right elbow point, RHP= Right
560   hand point, LEP= Left elbow point, LHP= Left hand point, MP= Mid-point,
561   LKP= left knee point, RKP= Right knee point, LFP= Left foot point,
562   RFP=Right foot point.
563   (Table 4) represents the results of multi-person human body parts for the
564   COCO dataset. For identified body parts, we indicated ✓ and for unidentified
565   body parts we adopted ✗. We a attained detection accuracy for human1 -
566   81.81%, human2 - 72.72%, human3 - 72.72%, human4- 72.72%, human5 -
567   72.72%, and the mean detection accuracy of 74.54%.

568

569   HP= Head point, NP= Neck point, REP= Right elbow point, RHP= Right hand
570   point, LEP= Left elbow point, LHP= Left hand point, MP= Mid-point, LKP=
571   left knee point, RKP= Right knee point, LFP= Left foot point, RFP=Right foot
572   point.
573

574   (Table 5) presents the results of multi-person human body parts for the pose
575   track dataset. For identified body parts, we use ✓ and for unidentified we
576   adopted use✗. The detection accuracies follow: for human1- 63.63%, human2
577   - 63.63%, human3 - 63.63%, human4 - 63.63%, human5 - 72.72% and the mean
578   detection accuracy is 65.45%.
579

580   HP= Head point, NP= Neck point, REP= Right elbow point, RHP= Right hand
581   point, LEP= Left elbow point, LHP= Left hand point, MP= Mid-point, LKP=
582   left knee point, RKP= Right knee point, LFP= Left foot point, RFP=Right foot
583   point.
584

585   Experiment II: Event Classification Accuracies
586   For gait event classification, we used a CNN-based deep learning approach.
587   The design method is evaluated by the Leave One Subject Out (LOSO) cross-
588   validation method. In Figure. 21, the results over the mpii-video-pose dataset
589   show 90.90% gait event classification and detection accuracy. After this, we
590   applied the deep belief network over the Olympic sports dataset and found
591   the stochastic remote sensing event classification results. Figure 22 represents
592   the confusion matrix for the COCO dataset with 89.09% mean accuracy for
593   gait event classification. Finally, CNN is applied over the pose track dataset,
594   with the mean gait event classification accuracy of 88.18%. Figure 23
595   represents the results in the shape of the confusion matrix for the pose track
596   dataset.
597

Figure 21. Confusion matrix results using CNN over Mpii-video-pose dataset.

598

599   BI=Bicycling, Ce=Conditioning exercise, Da= Dancing, Fh= Fishing and
600   hunting, Mp= Music playing, Ra= Religious activities, SP =Sports,
601   Tr=Transportation, Wi= Walking, Wa= Water activities, Wn=Winter
602   activities.
603

Figure 22. Confusion matrix results using CNN over the COCO dataset.

604   BI=Bicycling, Da= Dancing, Ce=Conditioning exercise, Fh= Fishing and
605   hunting, Ra= Religious activities, Mp= Music playing, SP =Sports, Wi=
606   Walking, Tr=Transportation, Wa= Water activities, Wn=Winter activities.
607
608

Figure 23. Confusion matrix results using CNN over the Pose track dataset

609
610
611
Ce=Conditioning exercise, BI=Bicycling, Da= Dancing, Mp= Music playing, Fh= Fishing and hunting, Ra= Religious activities, Tr=Transportation, SP =Sports, Wi= Walking, Wa= Water activities, Wn=Winter activities.

612

613   Experiment III: Comparison with Other Classification Algorithms

614
615
616
617
618
619
In this segment, we equate the recall, precision, and f-1 measure over the mpii-video-pose dataset, the COCO, and the posetrack dataset. For the classification of gait events we used Decision tree, Artificial Neural Network and we associated the consequences with the CNN. (Table 6) shows the results over the mpii-video-pose dataset, (Table 7) shows the results over the COCO dataset, and (Table 8) shows the results over the posetrack dataset.

620

621
622
623
624
BI=Bicycling, Ce=Conditioning exercise, Da= Dancing, Fh= Fishing and hunting, Mp= Music playing, Ra= Religious activities, SP =Sports, Tr=Transportation, Wi= Walking, Wa= Water activities, Wn=Winter activities.

625

626

627
628
629
BI=Bicycling, Da= Dancing, Ce=Conditioning exercise, Fh= Fishing and hunting, Ra= Religious activities, Mp= Music playing, SP =Sports, Wi= Walking, Tr=Transportation, Wa= Water activities, Wn=Winter activities.

630

631
632
633
Ce=Conditioning exercise, BI=Bicycling, Da= Dancing, Mp= Music playing, Fh= Fishing and hunting, Ra= Religious activities, Tr=Transportation, SP =Sports, Wi= Walking, Wa= Water activities, Wn=Winter activities.

634

635
636
Experiment IV: Performance analysis and comprehensive analysis of features selection

637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
For gait event classification, contextual features were proposed in this article. Four features are based upon the full human body that are DOF, periodic motion, non-periodic motion, and motion direction flow. One is based upon human body joints which are the rotational angular joints. Using these features we make a complete features vector to classify gait event detection with the help of data mining technique and CNN-based classification approach. To check the performance of our contextual features we adopted one more experiment in which various combinations of features are utilized to check the best combination. Initially, the combination of degree of freedom (DOF), periodic motion, and nonperiodic motion are used. After that, the combination of degree of freedom (DOF), periodic motion, nonperiodic motion, and motion direction flow, finally the combination of degree of freedom (DOF), periodic motion, nonperiodic motion, motion direction flow, and the rotational angular joint are used. The table 9 shows the overview of performance analysis. The complete combination shows more accurate results.

654 DOF= degree of freedom, PM = periodic motion, NM= nonperiodic motion,
655 MDF= motion direction flow, RAJ= rotational angular joint

656 Experiment V: Comparison of our Proposed System with State-of-
657 the-Art Techniques

658 *Fan et al. (2015)* developed a unique approach to estimate the human
659 pose which is based on deep learning-based dual-source CNN. As the input
660 they used patches of a given image and human body patches. After that they
661 combined both contextual and local index values to estimate human posture
662 with a better accuracy rate. *Pishchulin et al. (2016)* proposed a robust
663 formulation as a challenge of subsection partitioning and labeling (SPLP).
664 The SPLP structure, unlike previous two-stage methods that separated the
665 identification and pose estimation measures, suggests the number of persons,
666 certain poses, spatial proximity, including component level occlusions all at
667 the same time. In *Wei et al. (2016),* convolutional position devices have an
668 edge infrastructure for solving formal classifications based on computer
669 vision that does not require visual type reasoning. We demonstrated that by
670 transmitting increasingly refined confusion beliefs among points, a
671 sequential framework consisting of convolutional networks is incapable of
672 effectively training a structural component for the position. *Jin et al. (2019)*
673 proposed SpatialNet and TemporalNet combined to form a single pose
674 prediction and monitoring conceptual model: Body part identification and
675 part-level temporal classification are handled by SpatialNet, while the
676 contextual classification of human events is handled by TemporalNet. *Bao et*
677 *al. (2020)* suggest a hand gesture identification-by-tracking system that
678 incorporates pose input into both the video human identification and human
679 connection levels. A person's position prediction with pose descriptive
680 statistics is used in the first level to reduce the impact of distracting and
681 incomplete human identification in images. *Umer et al. (2020)* present a
682 method for detecting people in video which depends on key feature
683 connections. Rather than training the system to estimate key-point
684 communications on video sequences, the system is equipped to estimate
685 human pose utilizing personality on massive scale datasets. *Sun et al. (2018)*
686 suggested a technique for features extracted in which they remove guided
687 optical flow and use a CNN-based paradigm to identify and classify human
688 events. *Rachmadi, Uchimura & Koutaki (2016)* described a method for
689 dealing with event recognition and prevention using CNN and NNA
690 (Network in Network Architecture) frameworks, which are the foundation of
691 modern CNN. CNN's streamlined infrastructure, median, average, and
692 commodity features are used to define human activities. *Zhu et al. (2019)*
693 provide a detailed method for identifying incidents in security video.
694 Throughout the TRECVID-SED 2016 test, their method outperformed others
695 by a substantial margin by combining path modeling with deep learning.

696   (Table 9) shows the gait event mean accuracy comparison with the other
697   methods over the MPII, COCO and Pose track datasets.

698

699

700   **SCOPE OF THE PAPER:**

701   This article is based on human motion analysis and gait event detection-
702   based approaches, which are parts of computer vision, image processing,
703   data science, machine learning, deep learning, neural networks, and artificial
704   intelligence. Especially we can deploy this research project over airport
705   security, railways station, seaports, bus stations, metro stations, real-time
706   smart system environment and other smart surveillance systems. With
707   complex human-based crowed data which is the input of the proposed
708   system, human motion analysis and gait event classification are the
709   challenging tasks, with the help of various techniques, features, and
710   classification method. We achieve this with a much better accuracy rate than
711   previous methods. This system works fine in less crowed based datasets and
712   also on complex video data systems, but sometimes accuracy may be
713   compromised due to various factors such as distance, location, shadow, and
714   other illusion-based factors. This system also works on the real-time smart
715   system to analyze human motion and gait event detection.

716   **FAILURE CASES AND LIMITATIONS:**

717   This article is based on human motion analysis and gait event detection-
718   based approaches, which are parts of computer vision, image processing,
719   data science, machine learning, deep learning, neural networks, and artificial
720   intelligence. Especially we can deploy this research project over airport
721   security, railways station, seaports, bus stations, metro stations, real-time
722   smart system environment and other smart surveillance systems. With
723   complex human-based crowed data which is the input of the proposed
724   system, human motion analysis and gait event classification are the
725   challenging tasks, with the help of various techniques, features, and
726   classification method. We achieve this with a much better accuracy rate than
727   previous methods. This system works fine in less crowed based datasets and
728   also on complex video data systems, but sometimes accuracy may be
729   compromised due to various factors such as distance, location, shadow, and
730   other illusion-based factors. This system also works on the real-time smart
731   system to analyze human motion and gait event detection. Figure 24 shows
732   the detailed overview of failures and limitations of the proposed method in
733   which we can see the overlapping issue and complex scenarios to find human
734   and human landmarks.

735

Figure 24. Some examples of limitations and failure cases

736

737 Openpose *Viswakumar et al., (2019)* is a CNN-based approach that is used
738 for the human motion estimation and the gait analysis in various domains. We
739 applied openpsoe CNN classification and detected the human body parts. The
740 mean accuracy for the MPII dataset was 84.90%, the one for COCO dataset
741 was 86.00% and the one for Posetrack dataset was 85.18%. Table 11 shows
742 the complete results of openpose CNN.

743 HP= Head point, NP= Neck point, REP= Right elbow point, RHP= Right
744 hand point, LEP= Left elbow point, LHP= Left hand point, MP= Mid-point,
745 LKP= left knee point, RKP= Right knee point, LFP= Left foot point,
746 RFP=Right foot point.

747 Table 12 shows the detailed comparison of the openpose CNN model with
748 the proposed method. Results show the better accuracy of our proposed
749 method than the accuracy of the openpose method.
750

751 Figure 25 shows the detailed results of openpose CNN over human
752 sillhouttes.

Figure 25. Results of the openpose CNN based classification

753
754 **CONCLUSION**

755 This article is based on a reconstituted 3D synthetic model of the human body,
756 gait event detection and classification over complex human video articulated
757 datasets. Three benchmark datasets were selected for experiments: mpii-
758 video-pose, COCO, and pose tracking datasets. Initially, human detection and
759 landmark recognition are performed. After that, 2D human skeletons are
760 transformed into 3D synthetic-based models for the analysis of human
761 posture. For features reduction and optimization, a rule-based features mining
762 technique is adopted and finally, a deep learning classification algorithm
763 CNN is applied for gait event recognition and classification. For the mpii-
764 video pose dataset, we achieve the human landmark detection mean accuracy
765 of 87.09% and gait event recognition mean accuracy of 90.90%. For the
766 COCO dataset, we achieve the human landmark detection mean accuracy of
767 87.36 % and gait event recognition mean accuracy of 89.09%. For the pose
768 track dataset, we achieve the human landmark detection mean accuracy of
769 87.72% and gait event recognition mean accuracy of 88.18%. The proposed
770 system's performance shows a significant improvement compared to existing
771 state-of-the-art frameworks. The limitation of the proposed framework is due
772 to the complexity in the videos and group density which make it is difficult
773 to achieve more accurate results.

774 .
775

776 **Funding**
777 This research was supported by the Basic Science Research Program through the National Research
778 Foundation of Korea (NRF), funded by the Ministry of Education (No. 2018R1D1A1A02085645). Also,
779 this work was supported by the Korea Medical Device Development Fund grant funded by the Korea

780   government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of
781   Health & Welfare, the Ministry of Food and Drug Safety) (Project Number: 202012D05-02).
782

783   **Competing Interests**
784   The authors declare that they have no competing interests.
785

786   **Author Contributions**
787   • Yazeed Yasin Ghadi conceived and designed the experiments, authored or reviewed drafts of the
788   paper, and approved the final draft.
789   • Israr Akhter examined the experimental parameters, performed logical testing, analyzed the data
790   as well as computational work, prepared figures and/or tables, and approved the final draft.
791   • Mohammed Alarfaj analyzed the data, play vital role in major revision and approved the
792   final draft.
793   • Ahmad Jalal conceived and designed the experiments, figures and tables, authored or reviewed
794   drafts of the paper, and approved the final draft.
795   • Kibum Kim conceived and designed the experiments, authored or reviewed drafts of the paper,
796   and approved the final draft.
797

798   **Data Availability**
799   The following information was supplied regarding data availability.
800   Code and data are available in the Supplemental Files.
801

802   **Supplemental Information**
803   Supplemental information for this article can be found online at
804

805   **REFERENCES**

806   **Aggarwal JK, Cai Q. 1999**. Human Motion Analysis: A Review. *Computer Vision and Image*
807   *Understanding*. DOI: 10.1006/cviu.1998.0744.
808   **Akhter I. 2020**. Automated Posture Analysis of Gait Event Detection via a Hierarchical
809   Optimization Algorithm and Pseudo 2D Stick-Model. Ph.D. Thesis, Air University,
810   Islamabad, Pakistan, December 2020.
811   **Akhter I, Jalal A, Kim K**. Pose Estimation and Detection for Event Recognition using Sense-
812   Aware Features and Adaboost Classifier.
813   **Akhter I, Jalal A, Kim K. 2021**. Adaptive Pose Estimation for Gait Event Detection Using
814   Context-Aware Model and Hierarchical Optimization. *Journal of Electrical Engineering \&*
815   *Technology*:1–9.
816   **Amft O, Tröster G. 2008**. Recognition of dietary activity events using on-body sensors.
817   *Artificial Intelligence in Medicine*. DOI: 10.1016/j.artmed.2007.11.007.
818   **Bao Q, Liu W, Cheng Y, Zhou B, Mei T. 2020**. Pose-guided tracking-by-detection: Robust
819   multi-person pose tracking. *IEEE Transactions on Multimedia* 23:161–175.
820   **Cao Z, Simon T, Wei S-E, Sheikh Y. 2017**. Realtime multi-person 2d pose estimation using
821   part affinity fields. In: *Proceedings of the IEEE conference on computer vision and pattern*

822   *recognition*. 7291–7299.
823

824 **Einfalt M, Dampeyrou C, Zecha D, Lienhart R. 2019**. Frame-Level Event Detection in
825   Athletics Videos with Pose-Based Convolutional Sequence Networks. In: *Proceedings*
826   *Proceedings of the 2nd International Workshop on Multimedia Content Analysis in Sports -*
827   *MMSports '19*. New York, New York, USA: ACM Press, 42–50. DOI:
828   10.1145/3347318.3355525.
829 **Fan X, Zheng K, Lin Y, Wang S. 2015**. Combining local appearance and holistic view: Dual-
830   source deep neural networks for human pose estimation. In: *Proceedings of the IEEE*
831   *conference on computer vision and pattern recognition*. 1347–1355.
832 **Feng Q, Gao C, Wang L, Zhao Y, Song T, Li Q. 2020.** Spatio-temporal fall event detection in
833   complex scenes using attention guided LSTM. *Pattern Recognition Letters*. DOI:
834   10.1016/j.patrec.2018.08.031.
835 **Franco A, Magnani A, Maio D. 2020.** A multimodal approach for human activity recognition
836   based on skeleton and RGB data. *Pattern Recognition Letters*. DOI:
837   10.1016/j.patrec.2020.01.010.
838 **Franklin RJ, Mohana, Dabbagol V. 2020.** Anomaly Detection in Videos for Video
839   Surveillance Applications using Neural Networks. In: *Proceedings of the 4th International*
840   *Conference on Inventive Systems and Control, ICISC 2020*. DOI:
841   10.1109/ICISC47916.2020.9171212.
842 **Gochoo M, Akhter I, Jalal A, Kim K. 2021.** Stochastic Remote Sensing Event Classification
843   over Adaptive Posture Estimation via Multifused Data and Deep Belief Network. *Remote*
844   *Sensing* 13. DOI: 10.3390/rs13050912.
845 **Golestani N, Moghaddam M. 2020.** Human activity recognition using magnetic induction-
846   based motion signals and deep recurrent neural networks. *Nature Communications*. DOI:
847   10.1038/s41467-020-15086-2.
848 **Hassner T, Basri R. 2006.** Example based 3D reconstruction from single 2D images. In: *2006*
849   *Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*. 15.
850 **Hu Y, Jiang D, Yan S, Zhang L, others. 2004.** Automatic 3D reconstruction for face
851   recognition. In: *Sixth IEEE International Conference on Automatic Face and Gesture*
852   *Recognition, 2004. Proceedings*. 843–848.
853 **Jalal A, Akhtar I, Kim K. 2020**. Human Posture Estimation and Sustainable Events
854   Classification via Pseudo-2D Stick Model and K-ary Tree Hashing. *Sustainability* 12:9814.
855 **Jalal A, Khalid N, Kim K. 2020.** Automatic recognition of human interaction via hybrid
856   descriptors and maximum entropy markov model using depth sensors. *Entropy*. DOI:
857   10.3390/E22080817.
858 **Jiang YG, Dai Q, Mei T, Rui Y, Chang SF. 2015.** Super Fast Event Recognition in Internet
859   Videos. *IEEE Transactions on Multimedia*. DOI: 10.1109/TMM.2015.2436813.
860 **Jin S, Liu W, Ouyang W, Qian C. 2019.** Pose-guided tracking-by-detection: Robust multi-
861   person pose tracking In: *Proceedings of the IEEE/CVF Conference on Computer Vision and*
862   *Pattern Recognition (CVPR)*.
863 **Khalid N, Gochoo M, Jalal A, Kim K. 2021**. Modeling Two-Person Segmentation and
864   Locomotion for Stereoscopic Action Identification: A Sustainable Video Surveillance
865   System. *Sustainability* 13:970.
866 **Khan MA, Javed K, Khan SA, Saba T, Habib U, Khan JA, Abbasi AA. 2020**. Human action
867   recognition using fusion of multiview and deep features: an application to video

868   surveillance. *Multimedia Tools and Applications*. DOI: 10.1007/s11042-020-08806-9.

869 **Khan MH, Zöller M, Farid MS, Grzegorzek M. 2020**. Marker-based movement analysis of
870   human body parts in therapeutic procedure. *Sensors (Switzerland)*. DOI:
871   10.3390/s20113312.

872 **van der Kruk E, Reijne MM. 2018.** Accuracy of human motion capture systems for sport
873   applications; state-of-the-art review. *European Journal of Sport Science*. DOI:
874   10.1080/17461391.2018.1463397.

875 **Li A, Miao Z, Cen Y, Zhang X-P, Zhang L, Chen S. 2020.** Abnormal event detection in
876   surveillance videos based on low-rank and compact coefficient dictionary learning. *Pattern*
877   *Recognition* 108:107355.

878 **Liu J, Luo J, Shah M. 2009.** Recognizing realistic actions from videos in the Wild. In: *2009*
879   *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*
880   *Workshops, CVPR Workshops 2009*. DOI: 10.1109/CVPRW.2009.5206744.

881 **Lohithashva BH, Aradhya VNM, Guru DS. 2020.** Violent video event detection based on
882   integrated LBP and GLCM texture features. *Revue d'Intelligence Artificielle* 34:179–187.

883 **Mokhlespour Esfahani MI, Zobeiri O, Moshiri B, Narimani R, Mehravar M, Rashedi E,**
884   **Parnianpour M. 2017**. Trunk motion system (TMS) using printed body worn sensor
885   (BWS) via data fusion approach. *Sensors (Switzerland)*. DOI: 10.3390/s17010112.

886 **Pishchulin L, Insafutdinov E, Tang S, Andres B, Andriluka M, Gehler P V, Schiele B. 2016.**
887   Deepcut: Joint subset partition and labeling for multi person pose estimation. In:
888   *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4929–
889   4937.

890 **Rachmadi RF, Uchimura K, Koutaki G. 2016**. Combined convolutional neural network for
891   event recognition. In: *Proceedings of the Korea-Japan Joint Workshop on Frontiers of*
892   *Computer Vision*. 85–90.

893 **Sun S, Kuang Z, Sheng L, Ouyang W, Zhang W. 2018**. Optical Flow Guided Feature: A Fast
894   and Robust Motion Representation for Video Action Recognition. In: *Proceedings of the*
895   *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. DOI:
896   10.1109/CVPR.2018.00151.

897 **Tahir SB. 2020**. A Triaxial Inertial Devices for Stochastic Life-Log Monitoring via Augmented-
898   Signal and a Hierarchical Recognizer. Ph.D. Thesis, Air University, Islamabad, Pakistan,
899   December 2020.

900 **Tahir SB, Jalal A, Kim K.** IMU Sensor based Automatic-Features Descriptor for Healthcare
901   Patient's daily life-log Recognition.

902 **Uddin MZ, Thang ND, Kim JT, Kim T-S. 2011.** Human activity recognition using body joint-
903   angle features and hidden Markov model. *Etri Journal* 33:569–579.

904 **Ullah A, Muhammad K, Haq IU, Baik SW. 2019**. Action recognition using optimized deep
905   autoencoder and CNN for surveillance data streams of non-stationary environments. *Future*
906   *Generation Computer Systems*. DOI: 10.1016/j.future.2019.01.029.

907 **Umer R, Doering A, Leibe B, Gall J. 2020.** Self-supervised keypoint correspondences for
908   multi-person pose estimation and tracking in videos. *arXiv preprint arXiv:2004.12652*.

909 **Viswakumar A, Rajagopalan V, Ray T, Parimi C. 2019**. Human gait analysis using
910   OpenPose. In: *2019 Fifth International Conference on Image Information Processing*
911   *(ICIIP)*. 310–314.

912 **Rehman MA, Raza H, Akhter I. 2018**. SECURITY ENHANCEMENT OF HILL CIPHER BY
913   USING NON-SQUARE MATRIX APPROACH. In: *Proceedings of the 4th international*

914   *conference on knowledge and innovation in Engineering, Science and Technology*.
915   Acavent,. DOI: 10.33422/4kiconf.2018.12.24.
916   **Wang Y, Du B, Shen Y, Wu K, Zhao G, Sun J, Wen H. 2019.** EV-gait: Event-based robust
917   gait recognition using dynamic vision sensors. In: *Proceedings of the IEEE Computer*
918   *Society Conference on Computer Vision and Pattern Recognition*. DOI:
919   10.1109/CVPR.2019.00652.
920   **Wang Y, Mori G. 2008.** Multiple tree models for occlusion and spatial constraints in human
921   pose estimation. In: *Lecture Notes in Computer Science (including subseries Lecture Notes*
922   *in Artificial Intelligence and Lecture Notes in Bioinformatics)*. DOI: 10.1007/978-3-540-
923   88690-7-53.
924   **Wei S-E, Ramakrishna V, Kanade T, Sheikh Y. 2016.** Multi-Person Articulated Tracking
925   With Spatial and Temporal Embeddings,Proceedings of the IEEE/CVF Conference on
926   Computer Vision . In: *2016 IEEE Conference on*.
927   **Yu J, Lei A, Hu Y. 2019**. Soccer video event detection based on deep learning. In: *Lecture*
928   *Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and*
929   *Lecture Notes in Bioinformatics)*. DOI: 10.1007/978-3-030-05716-9_31.
930   **Zheng Y, Gu S, Edelsbrunner H, Tomasi C, Benfey P. 2011.** Detailed reconstruction of 3D
931   plant root shape. In: *2011 International Conference on Computer Vision*. 2026–2033.
932   **Zhu Y, Zhou K, Wang M, Zhao Y, Zhao Z. 2019**. A comprehensive solution for detecting
933   events in complex surveillance videos. *Multimedia Tools and Applications*. DOI:
934   10.1007/s11042-018-6163-6.
935   **Zou Y, Shi Y, Shi D, Wang Y, Liang Y, Tian Y. 2020**. Adaptation-Oriented Feature Projection
936   for One-shot Action Recognition. *IEEE Transactions on Multimedia*. DOI:
937   10.1109/tmm.2020.2972128.
938

Figure 1

The proposed system model's structural design.

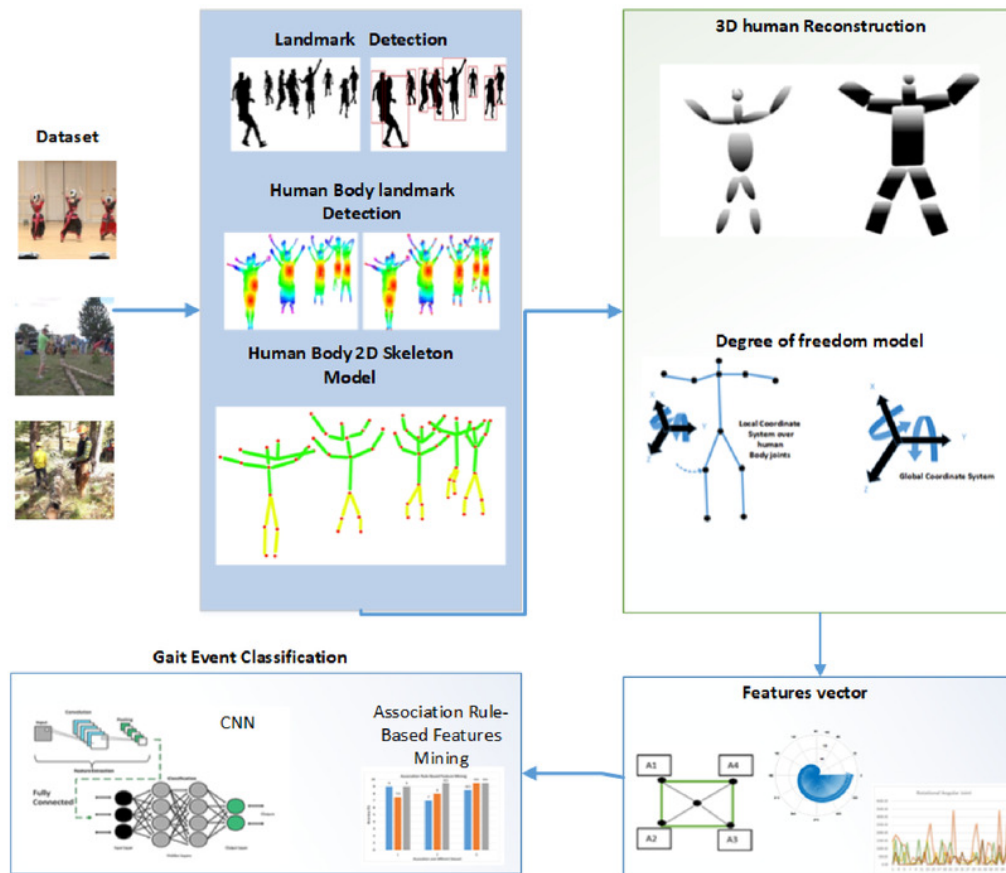



Figure 2

Flow chart of the proposed method.

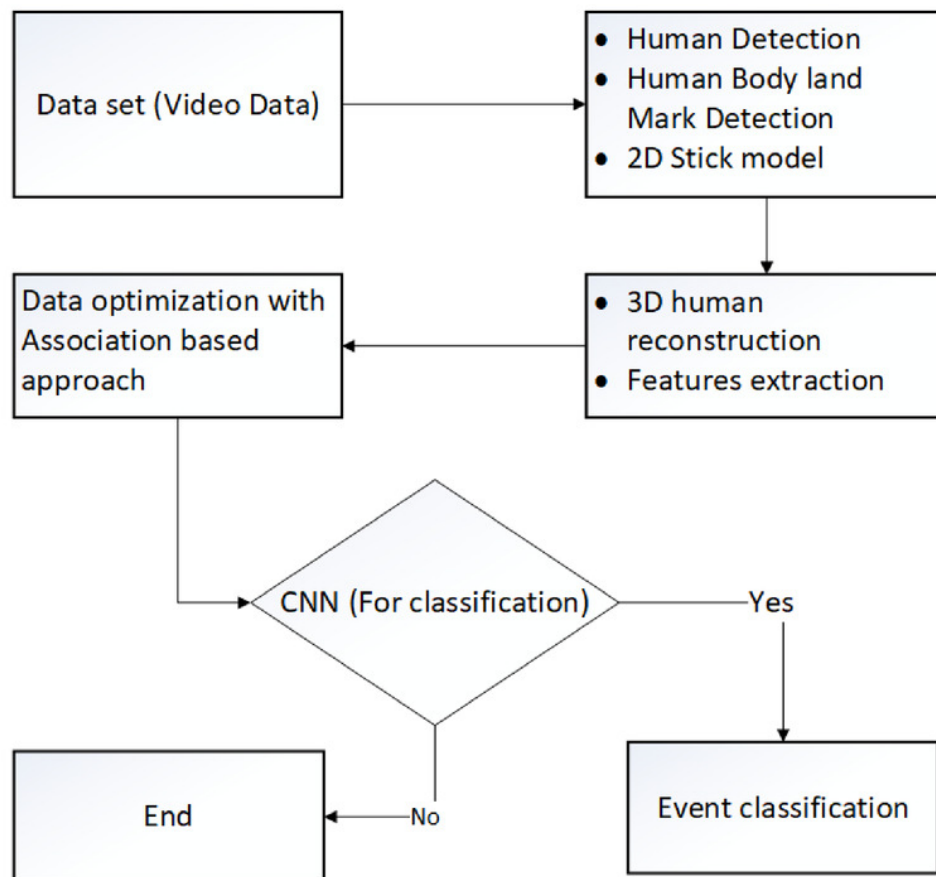



Figure 3

Results of different background subtraction techniques along with the original image.
(a) Original image (b) change detection (c) floor detection (d) Markov random field and
(e) spatial-temporal differencing.

(a) (b) (c)

(d) (e)

Figure 4

Results of (a) Optimized human silhouette (b) human head detection (c) human detection in RGB videos and image sequences.

Figure 5

Human body landmark detection results (a) presents the landmark results using an HSV color map, (b) presents the eleven human body points.

Figure 6

The Human 2D skeleton model results in over eleven human body parts.

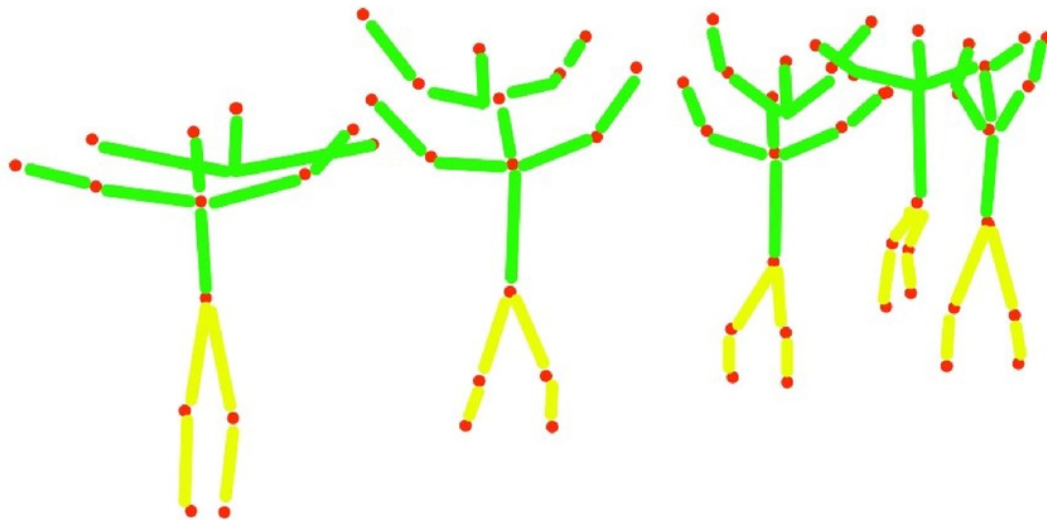



Figure 7

The results of the computational model with ellipsoids over human body points.

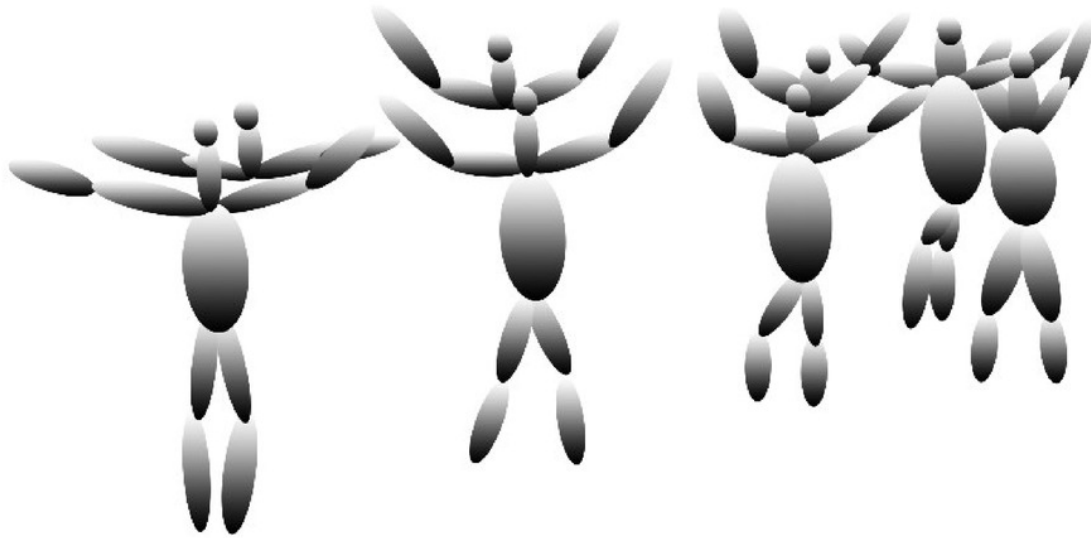
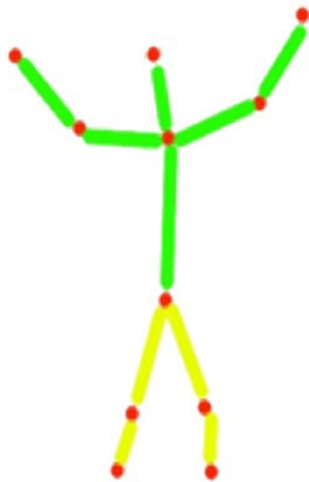



Figure 8

The results of the synthetic model with super quadrics over human body points. (a) Human 2D skeleton, (b) computational model with ellipsoids (c) synthetic model with super quadrics.

Figure 9

The theme concept of local and global coordinate systems. The left side shows the local coordinate system over the human left knee; the right side shows the DOF based global coordinate system.

Local Coordinate
System over
human
Body joints

Global Coordinate System

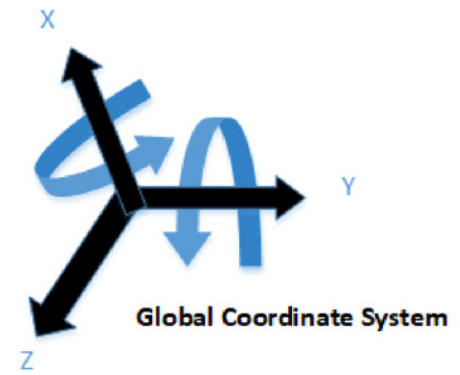
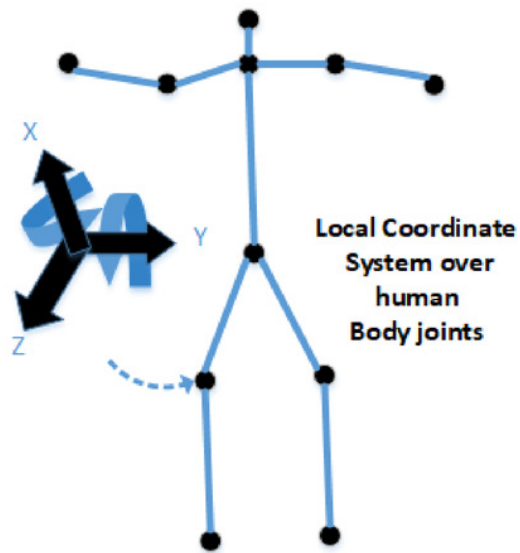



Figure 10

The results of the 3D ellipsoid reconstruction over the synthetic model with super quadrics and joint angle estimation

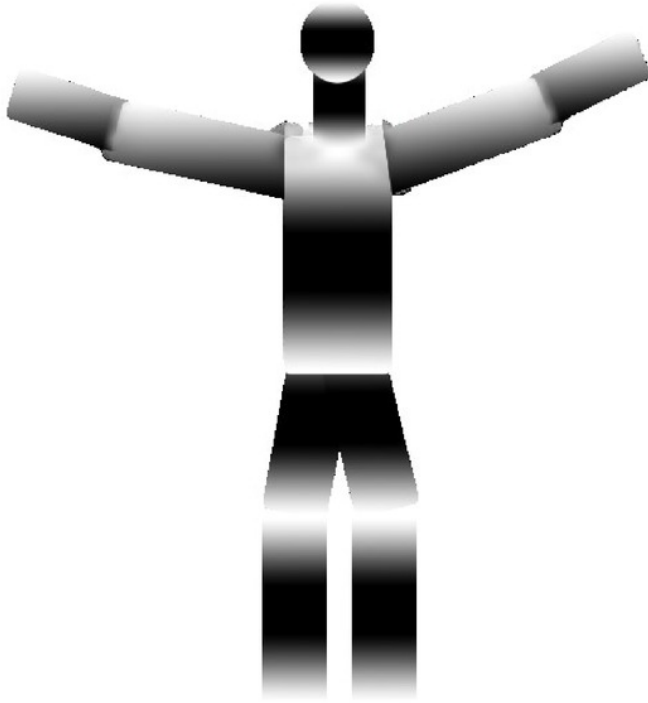



Figure 11

Few DOF results examples.

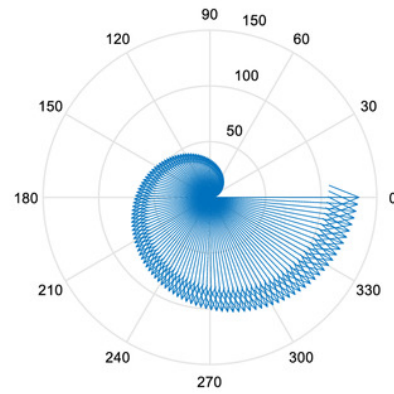
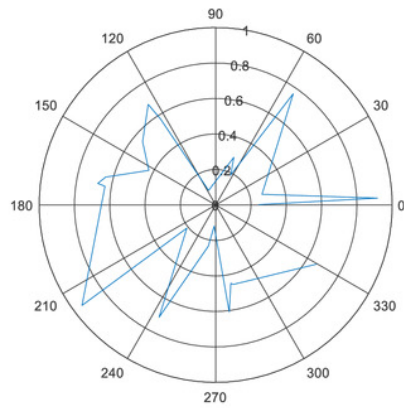



Figure 12

Periodic motion results.

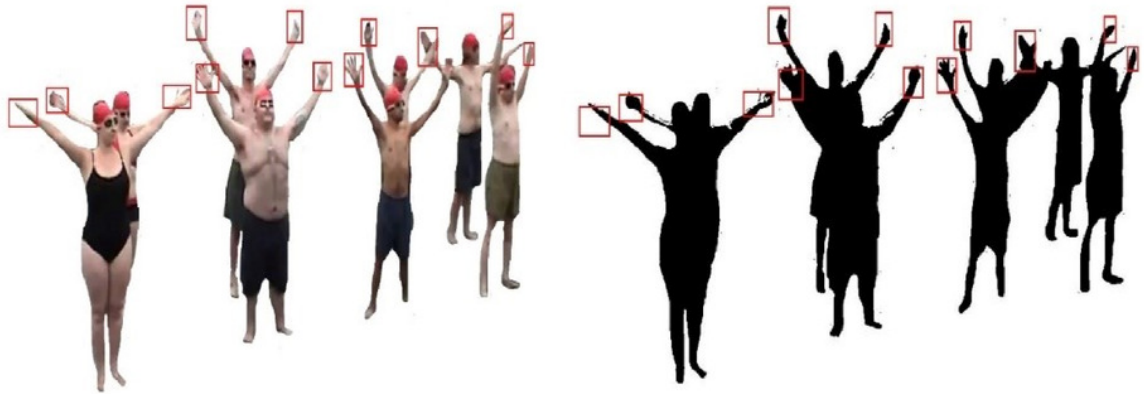



Figure 13

Results of non-periodic motion.

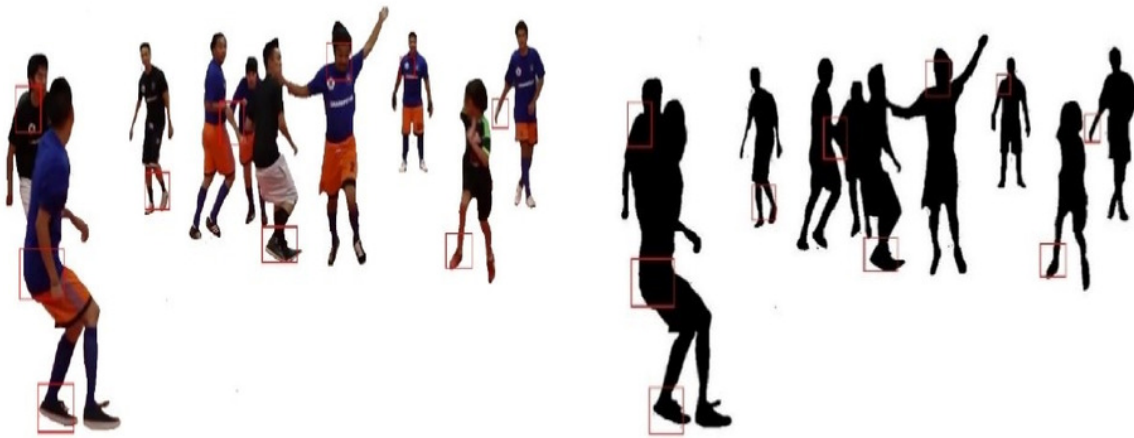



Figure 14

The results of motion direction flow over the basketball video.

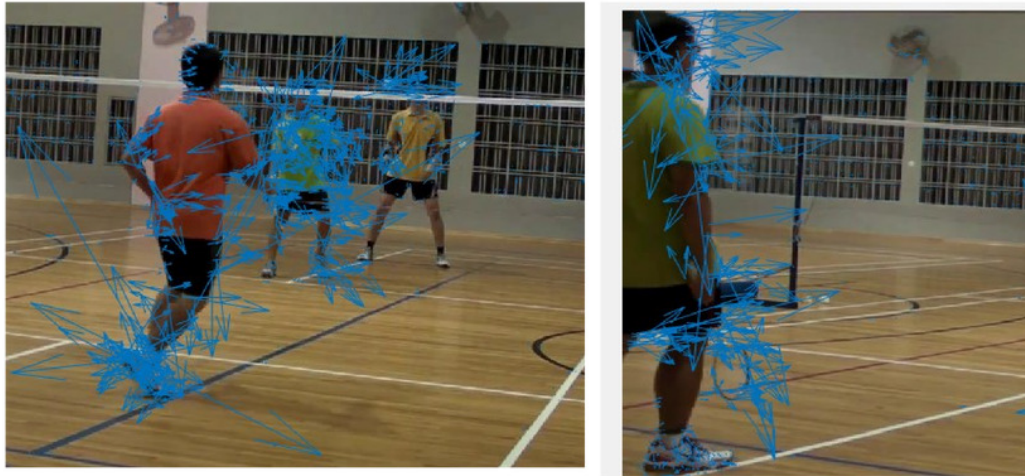



Figure 15

Rotational angular joint results and the pattern of rotational angels

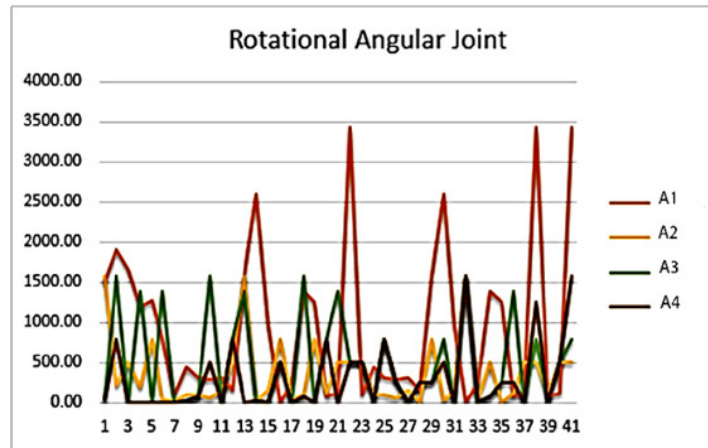
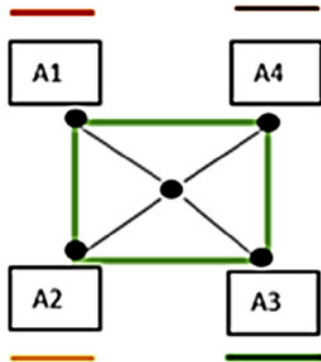



Figure 16

The most accurate features results via the rule-based features mining approach over the mpii-video pose, COCO, and Pose track datasets.

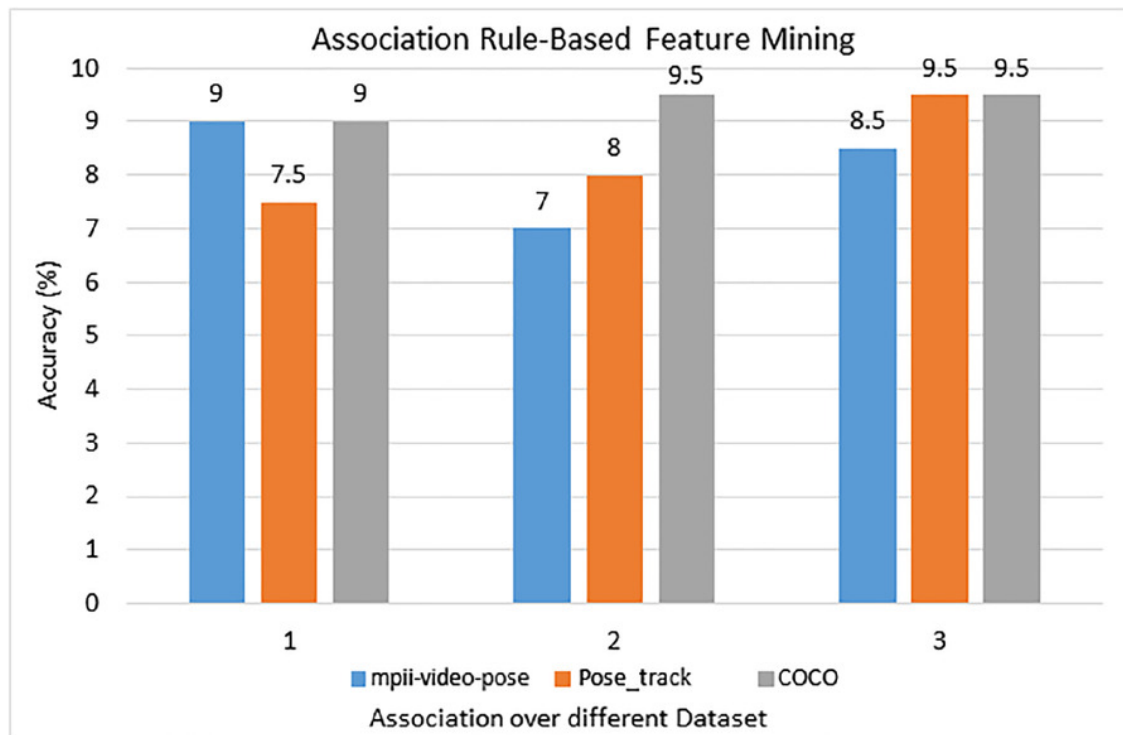



Figure 17

CNN model overview.

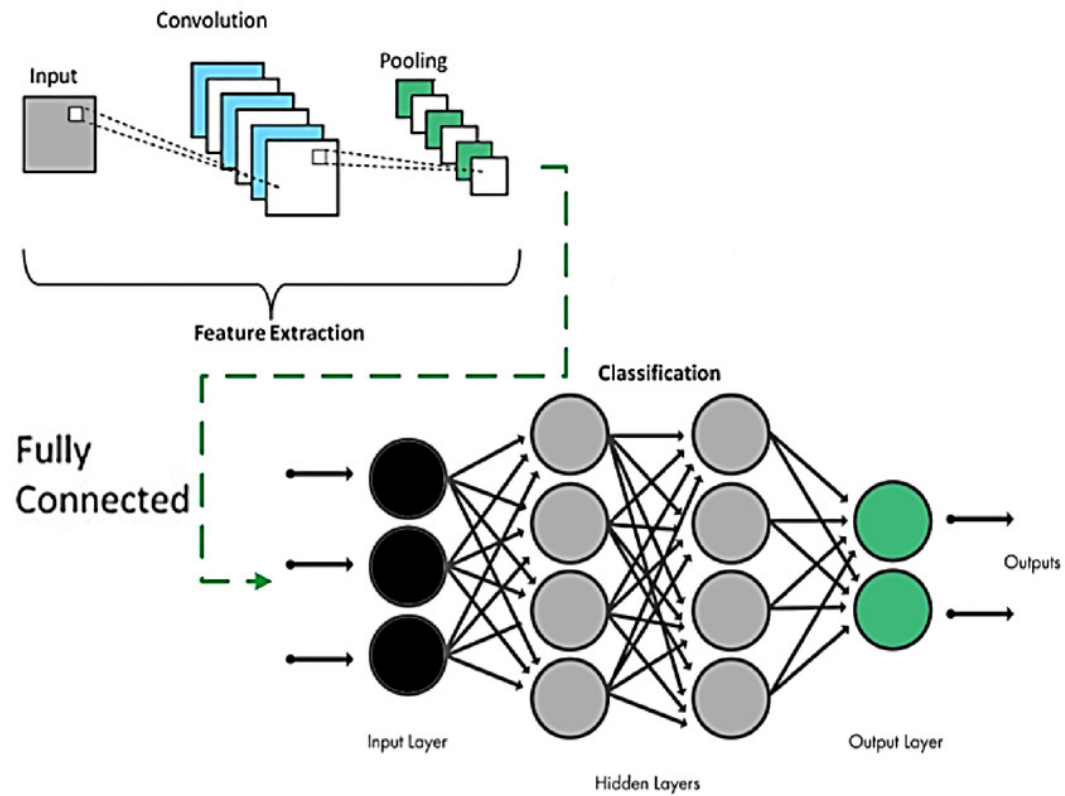



Figure 18

A few example images of the mpii-video-pose dataset.

Figure 19

A few example images from the COCO dataset.

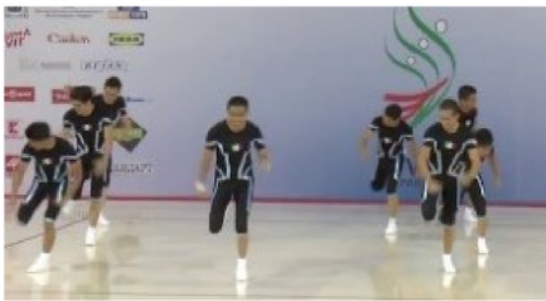



Figure 20

A few example images of the postmark dataset.

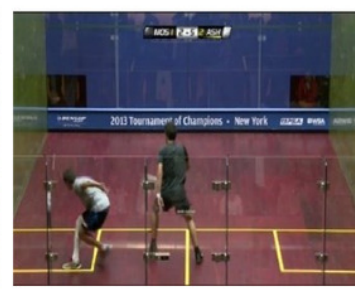



Figure 21

Confusion matrix results using CNN over Mpii-video-pose dataset.

	Bi	Ce	Da	Fh	Mp	Ra	Sp	Tr	Wl	Wa	Wn
Bi	9	1	0	0	0	0	0	0	0	0	0
Ce	0	9	0	0	0	0	1	0	0	0	0
Da	0	0	10	0	0	0	0	0	0	0	1
Fh	0	1	0	9	0	0	0	0	0	0	0
MP	1	0	0	0	9	0	0	0	0	0	0
Ra	0	0	1	0	0	9	0	0	0	0	0
SP	0	1	0	0	0	0	9	0	0	0	0
Tr	1	0	0	0	0	0	0	8	0	0	1
Wl	0	1	0	0	0	0	0	0	9	0	1
Wa	0	0	0	0	0	0	0	0	0	10	0
Wn	0	0	0	0	1	0	0	0	0	0	9

Gait event detection and classification mean accuracy = 90.90%

Figure 22

Confusion matrix results using CNN over the COCO dataset.

	Bi	Da	Ce	Fh	Ra	Mp	Sp	Wl	Tr	Wa	Wn
Bi	8	0	0	0	1	0	0	0	1	0	0
Da	0	10	0	0	0	0	0	0	0	0	0
Ce	0	0	9	0	1	0	0	0	0	0	0
Fh	0	0	0	8	1	1	0	0	0	0	0
Ra	0	0	0	0	10	0	0	0	0	0	0
Mp	0	0	1	0	0	8	0	0	1	0	0
Sp	1	0	0	0	0	0	9	0	0	0	0
Wl	0	0	0	0	1	0	0	9	0	0	0
Tr	0	1	0	0	0	0	0	0	8	0	1
Wa	0	0	1	0	0	0	0	0	0	9	0
Wn	0	0	0	0	0	0	0	0	0	0	10

Gait event detection and classification mean accuracy = 89.09%

Figure 23

Confusion matrix results using CNN over the Pose track dataset

	Ce	Bi	Da	Mp	Fh	Ra	Tr	Sp	Wl	Wn	Wa
Ce	7	0	0	1	0	0	1	0	1	0	0
Bi	0	8	1	0	0	0	0	0	0	1	0
Da	0	0	9	0	0	0	0	0	1	0	0
Mp	0	0	0	10	0	0	0	0	0	0	0
Fh	0	0	0	0	10	0	0	0	0	0	0
Ra	0	0	0	0	0	10	0	0	0	0	0
Tr	0	0	0	0	1	0	8	0	0	1	0
Sp	0	0	0	1	0	0	0	9	0	0	0
Wl	0	0	0	0	0	0	0	0	10	0	0
Wn	0	0	0	0	1	1	0	0	0	8	0
Wa	0	0	0	0	1	0	0	0	0	1	8

Gait event detection and classification mean accuracy = 88.18%

Figure 24

Some examples of limitations and failure cases

Figure 25

Results of the openpose CNN based classification

Algo01

Algorithm 1: Overview of the proposed method

Step 01.   **Input (Video based data set) Iv**

         Fun(Input)

         Video_data1, Video_data2, Video_data3

Step 02.   **Preprocessing on video data Pd**

         Fun(Nosie_redu)

         Fun(Frame_conv)

         Fun(Gry_scale, Bin_Covn)

    a.   Noise Reeducation (Nr)

    b.   Frame conversion (Fc)

    c.   Grey scale and binary conversion(Scb)

Step 03.   **Human body land mark detection()**

         Fun(Hum_Det)

         Fun(Landmark_det)

$$T_{He}^{q} \leftarrow T_{He}^{q-1} + \Delta T_{He}^{q-1}$$

         Fun(Human_2D_Stick_model)

$$Ubph = h \bowtie nq \bowtie S\_e \bowtie S\_hn$$
$$Mph = S\_t \bowtie Ubph$$

$$HLbs = hk \bowtie f \bowtie Mph$$

         Fun(3D_Human_recons)

$$C_{me} = P_a(e_x, e_y) \blacksquare P_{a+1}(e_x, e_y$$

    a.   Human Detection (Hd)

    b.   Landmark detection(Ld)

    c.   Human 2D stick model(Hsm)

    d.   3D Human reconstruction(Hr)

Step 04.   **Features extraction()**

         Fun(Peri_Motion)

$$PM(t) = \alpha \sin(\omega t + k)$$

         Fun(Non_Peri_Motion)

$$NPm(t) = \| \overline{P_{t,t+1}} - P_{t,t+2} \|$$

         Fun(Motion_Direction_Flow)

$$M_{df} = \sum_{0}^{p} I_{vl}(I) \rightarrow D$$

         Fun(Rotational_Angular_Joint)

$$A1 = \cos(x,y) \rightarrow L, \quad A2 = \cos(x,y) \rightarrow L$$
$$A3 = \cos(x,y) \rightarrow L, \quad A4 = \cos(x,y) \rightarrow L$$

    a.   PeriodicMotion(Pm)

    b.   NonPeriodicMotion(mn)

    c.   MotionDirectionandFlow(Mdf)

    d.   RotationalAngularJoint(Raj)

Step 05.   **Data optimization(Association based approach)**

$$nf_{(x,y)} = (g_x - g_y)\left(\frac{\Sigma_x - \Sigma_y}{2}\right)(g_x - g_y)^t$$

Step 06.   **Event classification()**

CNN Based event classification approach ()

$$TR_p = \sum_q wi_{p.q} \times a_p + b_q$$

Step 07.   **Output(Event Classification) Oc**

algo02

Algorithm 2: **Human Silhouette Optimization**

Input: EHS: Extracted Human Silhouettes
Output: Optimized human silhouette
/* human body localization in input data*/
/* WP is for white area*/
/* OS is optimized human silhouette*/
/* SF is denoting shape feature*/

Step 1:
Repeat
    For k=1 to I **do**
        For k=1 to I **do**
            $search$(WP)
        End
    End
    If WP1 > WP
        WP = WP1
    End
Until largest object shape searched in given frame.

Step 2:
/* Compare both WP */
For all pixel in both WP
    If $\text{WP}_{pixel\ information\ of\ frame\ 1} = \text{WP}_{information\ of\ frame\ 2}$
        $\text{WP}_{pixel\ information\ of\ frame\ 3} = \text{WP}_{pixel\ information\ of\ frame\ 1}$
    End
    If WP is inadequate for all inputs
        If pixel information is equal with SP
            $OS = \text{WP}_{pixel}$
        End
    End
End

Contextual Features Extraction

Algorithm 3: **Contextual Features Extraction**
Input: HS: Human silhouette from RGB video data
Output: Contextual feature vectors($Cf_1, Cf_2, Cf_3, ..., Cf_n$)
% feature vector for %
Contextual_features_vec ←[]
CF_vecsize ← GetFeaturesVectorsize ()
% loop over human silhouettes %
For i = 1:K
Contextual_features_vec _ interactions ← Get_ Contextual_features_vec (interactions)
        % extracting DOF, periodic motion, non periodic motion, motion direction and flow,
Rotational angular joint%
DOF ←ExtractDOF(Contextual_features_vec _ interactions)
PeriodicMotion←ExtractPeriodicMotion (Contextual_features_vec _ interactions)
NonPeriodicMotion ← ExtractNonPeriodicMotion (Contextual_features_vec _ interactions)
MotionDirectionandFlow←ExtractMotionDirectionandFlow(Contextual_features_vec_interactions)
RotationalAngularJoint←ExtractRotationalAngularJoint(Contextual_features_vec _ interactions)
Contextualvectors ← GetCFeaturevector
FVectors.append (CF_vectors)
End
Contextualvectors ← Normalize (Contextual_features_vec)

return Contextual_features_vec ($Cf_1, Cf_2, Cf_3, ..., Cf_n$)

Comprehensive review of relevant research.

Table 1 Comprehensive review of relevant research.

Human 2D posture analysis and event detection

Methods	Main contributions
Liu, Luo & Shah (2009)	Using contextual, stationary, and vibration attributes, an effective randomized forest-based methodology for human body part localization was developed. They used videos and photographs to evaluate different human actions.
Khan et al. (2020)	A micro, horizontal, and vertical differential function was proposed as part of an automated procedure. To classify human behavior, they used Deep Neural Network (DNN) mutation. To accomplish DNN-based feature strategies, a pre-trained Convolutional Neural Network Convolution layer was used.
Zou et al. (2020)	Adaptation-Oriented Features (AOF), an integrated framework with one-shot image classification for approximation to human actions was defined. The system applies to all classes, and they incorporated AOF parameters for enhanced performance.
Franco, Magnani & Maio (2020)	They created a multilayer structure with significant human skeleton details using RGB images. They used Histogram of Oriented (HOG) descriptor attributes to identify human actions.
Ullah et al. (2019)	The defined a single Convolutional Neural Network (CNN)-based actual data communications and information channel method. They utilized vision methods to gather information through non-monitoring instruments. The Convolutional Neural Network (CNN) technique is used to predict temporal features as well as deep auto-encoders and deep features in order to monitor human behavior.
van der Kruk & Reijne (2018)	They developed an integrated approach to calculate vibrant human motion in sports events using movement tracker sensors. The major contribution is the computation of human events in sports datasets by estimating the kinematics of human body joints, motion, velocity, and recreation of the human pose.
Wang & Mori (2008)	They developed a lightweight event recognition strategy based on spatial development and social body pose. The kinematics knowledge of attached human body parts is used to characterize tree-based characteristics.
Amft & Tröster (2008)	Using a Hidden Markov methodology, they built a solid framework for event identification which is accomplished using time-continuous dependent features and body marker detectors.
Wang et al. (2019)	With the assistance of a human tracking methodology, they developed a comprehensive new approach for estimating the accuracy of human motion. The Deep Neural Network (DNN) is used to identify events.
Jiang et al. (2015)	They introduced a multidimensional function method for estimating human motion and gestures. They used a late mean combination algorithm to recognize

	events in complex scenes.
Li et al. (2020)	They developed a lightweight organizational approach focused on optimal allocation, optical flow, and a histogram of the extracted optical flow. They were able to achieve effective event recognition using the standard optimization process, body joint restoration, and a Reduced and Compressed Coefficient Dictionary Learning (LRCCDL) methodology.
Einfalt et al. (2019)	Through task identification, isolation of sequential 2D posture characteristics and a convolutional sequence network, a coherent framework for event recognition with athletes in motion was created. They correctly identified number of sporting event.
Yu, Lei & Hu (2019)	Their work describes a probabilistic framework for detecting events in specific interchanges in soccer rivalry videos. This is done using the replay recognition approach which recognizes the most important background features for fulfilling spectator needs and generating replay storytelling clips.
Franklin, Mohana & Dabbagol (2020)	A comprehensive deep learning framework for identifying anomalous and natural events was developed. The findings were obtained using differentiation, grouping, and graph-based techniques. They discovered natural and unusual features for event duration use using deep learning techniques.
Cao et. Al (2019)	This article is based on a real-time method for detecting the 2D posture of numerous individuals in a picture. The suggested technique learns to connect body parts with persons in the image using a nonparametric informed decision-making Part Affinity Fields (PAFs).
3D human posture analysis and event detection	
Aggarwal & Cai (1999)	They devised a reliable method for analyzing the movement of human body parts through multiple cameras which monitor the body parts detection. They also created a simulation for human body joints that is 2D-3D.
Hassner & Basri (2006)	They designed an example-based synthesis methodology using a single class-based objects database that holds example reinforcements of realistic mappings due to the complexity of the objects.
Hu et al. (2004)	To define facial dimensionality, an effective 2D-to-3D hybrid face reconstruction technique is used to recreate a customizable 3D face template from a single cortical face picture with a neutral expression and regular lighting. Immersive-looking faces including different PIE are synthesized based on the customizable 3D image.
Zheng et al. (2011)	To enhance the classification of both the roots from each 2D image, they initially model the context only as a harmonic function. Second, they analyze the

	formalized graphical hull definition, which eliminates jitter and diffusion by maintaining continuity with a single 2D image. Third, they maintain connectivity by making variations to the 3D reconstruction by global errors minimization.
Uddin et al. (2011)	They proposed a heuristic approach for human activity detection and human posture analysis. For this, they utilized human body joint angle information with the help of the hidden Markov model (HMM).
Lohithashva, Aradhya & Guru (2020)	The researchers created a deep learning system for detecting abnormal and normal events. Distinction, classification, and graph-based methods were used to obtain the results. Using deep learning methods, they explored natural and uncommon features for event interval use.
Feng et al. (2020)	To retrieve deep features' spatial locations in composite images, a guided Long Short-Term Memory (LSTM) approach that is based on a Convolutional Neural Network (CCN) system was evaluated. For personal authentication, the state-of-the-art YOLO v3 template was used and, for event recognition, a directed Long Short-Term Memory (LSTM) driven method was used.
Khan et al. (2020)	They developed home-based patient control strategies based on body-marker detectors. To record data from patients, body-marker sensors with a color indicator framework are connected to the joints.
Mokhlespour Esfahani et al. (2017)	For sporting events, human movement monitoring body-marker tools were used to establish a Trunk Motion Method (TMM) with Body-worn Sensors that provide a low power physical system (BWS). Twelve removable detectors were used to measure 3D trunk movements in this process.
Golestani & Moghaddam,(2020)	A robust wireless strategy was developed for detecting physical human behavior. They used a magnetic flux cable to monitor human behavior, and thematic maps were attached to the body joints. Research lab approximation function and Deep RNN (Recurrent Neural Network) were used to enhance efficiency.

Human body parts recognition and detection accuracy

Table 2: **Human body parts recognition and detection accuracy**

Body key points	Distance	MPII (%)	Distance	COCO (%)	Distance	Posetrack (%)
HP	11.2	88	9.70	88	9.90	91
NP	10.8	86	10.2	86	11.1	88
REP	11.5	82	10.1	83	14.1	86
RHP	12.1	81	11.7	82	12.7	83
LEP	11.1	83	11.9	79	11.0	88
LHP	12.0	77	11.7	81	12.0	79
MP	10.1	91	13.1	90	11.9	91
LKP	13.2	94	12.8	92	12.3	87
RKP	9.90	91	10.3	91	11.7	81
LFP	10.3	94	11.2	95	14.1	94
RFP	11.5	91	10.3	94	13.8	97
Mean Accuracy Rate		**87.09**		**87.36**		**87.72**

Human body parts results of multi-person for mpii-video-pose dataset

Body parts	Human1	Human2	Human3	Human4	Human5
Table 3: Human body parts results of multi-person for mpii-video-pose dataset					
HP	✓	✓	✓	✓	✗
NP	✗	✓	✗	✗	✓
REP	✓	✓	✓	✓	✓
RHP	✓	✗	✗	✓	✓
LEP	✗	✓	✓	✓	✗
LHP	✗	✓	✓	✗	✓
MP	✓	✗	✓	✓	✓
LKP	✓	✓	✗	✗	✓
RKP	✗	✗	✓	✓	✗
LFP	✓	✓	✗	✓	✓
RFP	✓	✓	✓	✓	✓
Accuracy	63.63%	72.72%	63.63%	72.72%	72.72%
Mean accuracy = 69.09%					

Human body parts results of multi-person for COCO dataset

Table 4: Human body parts results of multi-person for COCO dataset					
Body parts	Human1	Human2	Human3	Human4	Human5
HP	✓	✓	✓	✓	✓
NP	✓	✓	✓	✗	✓
REP	✓	✗	✗	✓	✓
RHP	✓	✗	✗	✗	✗
LEP	✓	✓	✓	✓	✓
LHP	✗	✓	✓	✓	✗
MP	✗	✗	✓	✓	✓
LKP	✓	✓	✗	✗	✓
RKP	✓	✓	✓	✓	✗
LFP	✓	✓	✓	✓	✓
RFP	✓	✓	✓	✓	✓
Accuracy	81.81%	72.72%	72.72%	72.72%	72.72%
Mean accuracy = 74.54%					

: Human body parts results of multi-person for mpii-video-pose dataset

Body parts	Human1	Human2	Human3	Human4	Human5
HP	✓	✓	✓	✓	✓
NP	✗	✗	✓	✗	✓
REP	✓	✓	✗	✓	✓
RHP	✗	✗	✓	✗	✗
LEP	✓	✓	✗	✓	✓
LHP	✓	✗	✓	✗	✗
MP	✗	✓	✓	✓	✓
LKP	✓	✓	✗	✓	✓
RKP	✓	✗	✓	✗	✗
LFP	✓	✓	✗	✓	✓
RFP	✗	✓	✓	✓	✓
Accuracy	63.63%	63.63%	63.63%	63.63%	72.72.1%

Table 5: **Human body parts results of multi-person for mpii-video-pose dataset**

Mean accuracy = 65.45%

Precision, recall, and F-1 measure comparison with the artificial neural network, decision tree and CNN over Mpii-video-pose dataset

Table 6. Precision, recall, and F-1 measure comparison with the artificial neural network, decision tree and CNN over Mpii-video-pose dataset

Event Classes	Artificial Neural Network			Decision Tree			CNN		
Events	Precision	Recall	F- 1 measure	Precision	Recall	F- 1 measure	Precision	Recall	F- 1 Measure
Bi	0.778	0.700	0.737	0.667	0.600	0.632	0.818	0.900	0.857
Ce	0.700	0.700	0.700	0.700	0.700	0.700	0.692	0.900	0.783
Da	0.857	0.600	0.706	0.818	0.900	0.857	0.909	0.909	0.909
Fh	0.909	1.000	0.952	0.727	0.800	0.762	1.000	0.900	0.947
MP	0.900	0.900	0.900	0.727	0.800	0.762	0.900	0.900	0.900
Ra	0.889	0.800	0.842	0.889	0.800	0.842	1.000	0.900	0.947
SP	0.727	0.889	0.800	0.833	1.000	0.909	0.900	0.900	0.900
Tr	0.875	0.778	0.824	0.909	1.000	0.952	1.000	0.800	0.889
Wl	0.875	0.700	0.778	1.000	0.700	0.824	1.000	0.818	0.900
Wa	0.818	1.000	0.900	1.000	0.900	0.947	1.000	1.000	1.000
Wn	0.769	1.000	0.870	0.800	0.800	0.800	0.750	0.900	0.818
Mean	**0.827**	**0.824**	**0.819**	**0.825**	**0.818**	**0.817**	**0.906**	**0.893**	**0.896**

Precision, recall, and F-1 measure comparison with the artificial neural network, decision tree and CNN over COCO dataset.

Table 7. Precision, recall, and F-1 measure comparison with the artificial neural network, decision tree and CNN over COCO dataset.

Event Classes	Artificial Neural Network			Decision Tree			CNN		
Events	Precision	Recall	F-1 measure	Precision	Recall	F-1 measure	Precision	Recall	F-1 Measure
Bi	0.818	0.750	0.783	0.700	0.700	0.700	0.889	0.800	0.842
Da	0.750	0.750	0.750	0.889	0.800	0.842	0.909	1.000	0.952
Ce	0.889	0.667	0.762	0.833	1.000	0.909	0.818	0.900	0.857
Fh	0.900	1.000	0.947	0.818	0.900	0.857	1.000	0.800	0.889
Ra	0.909	0.909	0.909	0.818	0.900	0.857	0.714	1.000	0.833
Mp	0.900	0.818	0.857	0.900	0.900	0.900	0.889	0.800	0.842
Sp	0.667	0.857	0.750	0.889	0.800	0.842	1.000	0.900	0.947
Wl	0.900	0.818	0.857	0.727	1.000	0.842	1.000	0.900	0.947
Tr	0.900	0.750	0.818	0.889	0.800	0.842	0.800	0.800	0.800
Wa	0.800	1.000	0.889	1.000	0.800	0.889	1.000	0.900	0.947
Wn	0.727	1.000	0.842	0.875	0.700	0.778	0.909	1.000	0.952
Mean	0.833	0.847	0.833	0.849	0.845	0.842	0.903	0.891	0.892

Precision, recall, and F-1 measure comparison with the artificial neural network, decision tree and CNN over Posetrack dataset.

Table 8. Precision, recall, and F-1 measure comparison with the artificial neural network, decision tree and CNN over Posetrack dataset.

Event Classes	Artificial Neural Network			Decision Tree			CNN		
Events	Precision	Recall	F- 1 measure	Precision	Recall	F- 1 measure	Precision	Recall	F- 1 Measure
Ce	0.818	0.750	0.783	0.769	1.000	0.870	1.000	0.700	0.824
Bi	0.700	0.700	0.700	0.750	0.818	0.783	1.000	0.800	0.889
Da	0.889	0.667	0.762	0.875	0.700	0.778	0.900	0.900	0.900
Mp	0.900	1.000	0.947	0.778	0.700	0.737	0.833	1.000	0.909
Fh	0.875	0.875	0.875	0.692	0.900	0.783	0.769	1.000	0.870
Ra	0.900	0.818	0.857	0.818	0.900	0.857	0.909	1.000	0.952
Tr	0.750	0.900	0.818	0.700	0.700	0.700	0.889	0.800	0.842
Sp	0.857	0.750	0.800	0.875	0.700	0.778	1.000	0.900	0.947
Wl	0.857	0.667	0.750	1.000	0.750	0.857	0.833	1.000	0.909
Wn	0.800	1.000	0.889	1.000	1.000	1.000	0.727	0.800	0.762
Wa	0.769	1.000	0.870	0.778	0.778	0.778	1.000	0.800	0.889
Mean	0.829	0.830	0.823	0.821	0.813	0.811	0.896	0.882	0.881

Performance analysis over various features extracting upon the MPII, COCO and Pose track datasets.

Table9. Performance analysis over various features extracting upon the MPII, COCO and Pose track datasets.

Features combination	MPII (%)	COCO (%)	Pose Track (%)
DOF, PM, NPM	71.20	73.21	72.59
DOF, PM, NPM,MDF	76.73	76.07	75.13
DOF, PM, NPM,MDF,RAJ	**90.90**	**89.09**	**88.18**

Gait event mean accuracy comparison with the other methods over the MPII, COCO and Pose track datasets.

Table 10. Gait event mean accuracy comparison with the other methods over the MPII, COCO and Pose track datasets.

Methods	MPII (%)	Methods	COCO (%)	Methods	Pose Track (%)
Fan et al. (2015)	73.00	*Sun et al. (2018)*	74.20	*Jin et al. (2019)*	71.08
Pishchulin et al. (2016)	87.10	*Rachmadi, Uchimura & Koutaki (2016)*	82.30	*Bao et al. (2020)*	72.03
Wei et al. (2016)	90.50	*Zhu et al.(2019)*	83.10	*Umer et al. (2020)*	74.02
Proposed method	90.90		89.09		88.18

Human body parts recognition and detection accuracy using Openpose CNN

Table 11: **Human body parts recognition and detection accuracy using Openpose CNN**

Body key points	Distance	MPII (%)	Distance	COCO (%)	Distance	Posetrack (%)
HP	12.5	86	10.1	84	11.1	83
NP	12.3	83	12.3	83	12.5	81
REP	10.6	81	12.5	82	12.9	80
RHP	13.1	88	14.1	86	10.1	85
LEP	12.3	84	10.6	80	10.9	82
LHP	14.0	78	10.8	83	11.6	77
MP	11.3	88	12.2	88	9.9	92
LKP	11.0	85	11.9	87	10.5	89
RKP	15.1	83	13.3	89	12.8	85
LFP	14.2	90	11.6	91	11.3	91
RFP	12.3	88	11.1	93	10.1	92
Mean Accuracy Rate		**84.90**		**86.00**		**85.18**

Comparison table the Openpose CNN model with the proposed method

Table 12: **Comparison table the Openpose CNN model with the proposed method**

Dataset	Openpose CNN	Proposed method
MPII (%)	84.90	87.09
COCO (%)	86.00	87.36
Posetrack (%)	85.18	87.72