

Syntactic Model-Based Human Body 3D Reconstruction and Event Classification via Association based Features Mining and Deep Learning (#60580)

1

First submission

Guidance from your Editor

Please submit by **14 Jun 2021** for the benefit of the authors (and your \$200 publishing discount) .



Structure and Criteria

Please read the 'Structure and Criteria' page for general guidance.



Raw data check

Review the raw data.



Image check

Check that figures and images have not been inappropriately manipulated.

Privacy reminder: If uploading an annotated PDF, remove identifiable information to remain anonymous.

Files

Download and review all files from the [materials page](#).

22 Figure file(s)

2 Box file(s)

9 Table file(s)

3 Raw data file(s)



Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. BASIC REPORTING
2. EXPERIMENTAL DESIGN
3. VALIDITY OF THE FINDINGS
4. General comments
5. Confidential notes to the editor

 You can also annotate this PDF and upload it as part of your review

When ready [submit online](#).

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your [guidance page](#).





BASIC REPORTING

-  Clear, unambiguous, professional English language used throughout.
-  Intro & background to show context. Literature well referenced & relevant.
-  Structure conforms to [PeerJ standards](#), discipline norm, or improved for clarity.
-  Figures are relevant, high quality, well labelled & described.
-  Raw data supplied (see [PeerJ policy](#)).

EXPERIMENTAL DESIGN

-  Original primary research within [Scope of the journal](#).
-  Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
-  Rigorous investigation performed to a high technical & ethical standard.
-  Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

-  Impact and novelty not assessed. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
-  All underlying data have been provided; they are robust, statistically sound, & controlled.
-  Speculation is welcome, but should be identified as such.
-  Conclusions are well stated, linked to original research question & limited to supporting results.



The best reviewers use these techniques

Tip

Example

Support criticisms with evidence from the text or from other sources

Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.

Give specific suggestions on how to improve the manuscript

Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).

Comment on language and grammar issues

The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult. I suggest you have a colleague who is proficient in English and familiar with the subject matter review your manuscript, or contact a professional editing service.

Organize by importance of the issues, and number your points

1. Your most important issue
2. The next most important item
3. ...
4. The least important points

Please provide constructive criticism, and avoid personal opinions

I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC

Comment on strengths (as well as weaknesses) of the manuscript

I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.

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

Yazeed Yasin Ghadi¹, Israr Akhter², Munkhjargal Gochoo³, Ahmad Jalal², Kibum Kim^{Corresp. 4}

¹ Department of Computer Science and Software Engineering,, Al Ain University, Al Ain 15551, UAE, Al Ain, UAE, UAE, UAE

² Department of Computer Science,, Air University, 46000, Islamabad, Pakistan, Islamabad,, Pakistan, Pakistan

³ Department of Computer Science and Software Engineering,, United Arab Emirates University, Al Ain 15551, UAE, Al Ain,, UAE, UAE

⁴ Department of Human-Computer Interaction,, Hanyang University, Ansan,, South Korea, South Korea

Corresponding Author: Kibum Kim

Email address: kibum@hanyang.ac.kr

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, initially 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 human landmark detection mean accuracy of 87.72% 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

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

Yazeed Yasin Ghadi¹, Israr Akhter², Munkhjargal Gochoo³, Ahmad Jalal², and Kibum Kim⁴

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 Computer Science and Software Engineering, United Arab Emirates University, Al Ain 15551, UAE; mgochoo@uaeu.ac.ae

⁴ Department of Human-Computer Interaction, Hanyang University, Ansan, 15588, South Korea kibum@hanyang.ac.kr;

Corresponding Author:

Kibum Kim⁴

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, initially 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 human landmark detection mean accuracy of 87.72% 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, 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.

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, initially preprocessing the video samples for frame conversion, motion blur noise reduction and resizing. The next step in human detection to process these data sources by various algorithms for silhouette extraction, optimization of detected human silhouettes and human body landmark detection. The next stage is to analyze the human posture information. For this, 2D to 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 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 and optimization of humans and the extraction and optimization of human silhouettes is 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, contextual features, 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.

Table 1 Comprehensive review of relevant research.

Human 2D posture analysis and event detection

Methods	Main contributions
<i>Liu, Luo & Shah (2009)</i>	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.
<i>Khan et al. (2020)</i>	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.

<i>Zou et al. (2020)</i>	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.
<i>Franco, Magnani & Maio (2020)</i>	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.
<i>Ullah et al. (2019)</i>	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.
<i>van der Kruk & Reijne (2018)</i>	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.
<i>Wang & Mori (2008)</i>	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.
<i>Amft & Tröster (2008)</i>	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.
<i>Wang et al. (2019)</i>	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.
<i>Jiang et al. (2015)</i>	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.
<i>Li et al. (2020)</i>	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.
<i>Einfalt et al. (2019)</i>	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.
<i>Yu, Lei & Hu (2019)</i>	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.
<i>Franklin, Mohana & Dabbagol (2020)</i>	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.
3D human posture analysis and event detection	
<i>Aggarwal & Cai (1999)</i>	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.
<i>Hassner & Basri (2006)</i>	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.
<i>Hu et al. (2004)</i>	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.
<i>Zheng et al. (2011)</i>	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.
<i>Uddin et al. (2011)</i>	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).
<i>Lohithashva, Aradhya & Guru (2020)</i>	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.
<i>Feng et al. (2020)</i>	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.

MATERIALS & METHODS

For the video input of our proposed method, the main source is RGB cameras which provide clear information. The first step is video to frame conversion 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.

Preprocessing of the data

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

Background subtraction

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

Silhouette optimization and human detection

In this sub-phase of landmark detection, we find the optimized human silhouette through the merging of change detection, floor detection, Markov random field, and spatial-temporal differencing techniques with the help of an adaptive threshold approach. (Algorithm 1) shows the detailed procedure of silhouette optimization.

Algorithm 1: 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 WP_{pixel information of frame 1} = WP_{information of frame 2}

WP_{pixel information of frame 3} = WP_{pixel information of frame 1}

End

If WP is inadequate for all inputs

If pixel information is equal with SP

OS = WP_{pixel}

End

End

End

After this, human detection is performed in two phases, initially, head detection is performed with the help of a human head “size and shape-based” technique. We set the weight of a human head as $w_0=1/25$ of the human silhouette and, using region of interest model, we find the super pixel position of the human body and, after this, Gaussian kernel is used to capture the likely area of the human head. Finally, using this human head information, human shape and appearance information, human body movement, and motion information, human detection and identification is performed. Eq. (1) is used for head tracking

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

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

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

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.

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:

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

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

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

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;

$$T_t^q \leftarrow T_{to}^{q-1} + \Delta T_{to}^{q-1} \quad (5)$$

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;

$$T_k^q = (T_m^q - T_f^q) / 4 \quad (6)$$

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;

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

where T_e^q denotes the human elbow point, T_{hn}^q is the human hand point, and T_{nq}^q denotes the neck point. Figure 4 represents the results of landmarks and body parts.

In this section, basic 2D human body skeletonization is achieved, using extracted human body key points. Figure 5 represents the detailed view of the human body 2D skeleton Jalal, Akhtar & Kim (2020) over eleven human body parts. The human body 2D skeleton is based on three main body areas: Upper body parts of the human body $Ubph$, Mid parts of the human body Mph , and lower body parts of the human body $Lbph$. $Ubph$ denote the connection of the human head (h), neck point (nq), elbow (e), and human hand points (hn). Mph is initiated through the human midpoint (t) and $Unph$. $Lbph$ is based on human knee points (hk) and footpoints (f). Every human body part or joint takes a specific letter k to complete a particular action. Eq. (8), (9) and (10) demonstrate the calculated associations of the human 2D stick model as;

$$Ubph = h \bowtie nq \bowtie S_e \bowtie S_{hn} \quad (8)$$

$$Mph = S_t \bowtie Ubph \quad (9)$$

$$HLbs = hk \bowtie f \bowtie Mph \quad (10)$$

3D human reconstruction

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

A. Computational model with ellipsoids

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

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

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 6 shows 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.

$$S_{SQ} = C_{me} \rightarrow s_e \quad (12)$$

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 7 shows the results of the synthetic model with super quadrics over the computational model with ellipsoids.

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.

$$T_{Ne}^q \leftarrow T_{Ne}^{q-1} + \Delta T_{Ne}^{q-1} \quad (13)$$

where T_{Ne}^q represents the neck landmark location in any given video frame, q is consequential to calculation of the frame differences. See Fig. 8.

$$F_v = [\theta_{g_l}, \theta_{S_h}, \theta_{S_n}, \theta_{R_e}, \theta_{L_e}, \theta_{R_h}, \theta_{L_h}, \theta_{S_m}, \theta_{R_k}, \theta_{L_k}, \theta_{R_f}, \theta_{L_f}] \quad (14)$$

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.

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

Where $R_E(x,0 | I,D)$ 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 9 shows 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.

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

where D_{of} represents the degree of freedom feature vector, $\theta x, \theta y, \theta z$ shows the three dimension of the angle and εD is the local and global coordinate system. Figure 10 shows the results for the degree of freedom:

B. Periodic motion

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

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

where $PM(t)$ denotes periodic motion and $\alpha \sin(\omega t + k)$ shows the relation of human motion that is repeated in any given sequence of images. Figure 11 shows the results of periodic motion:

C. Non-periodic motion

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

$$Npm(t) = \| P_{t,t+1} - P_{t,t+2} \| \quad (18)$$

where Npm denotes non-periodic motion and $P_{t,t+1} - P_{t,t+2}$ shows the difference between the first and the next sequence of images in input data. Figure 12 shows the results of non-periodic motion:

D. Motion direction flow

For the identification of more accurate gait events motion direction flow is one a contributions in terms of contextual features. Using changes in motion 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.

$$M_{df} = \sum_0^p I_{vl}(I) \rightarrow D \quad (19)$$

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 13 shows the results of motion direction flow over the basketball class 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.

$$\begin{aligned} A1 &= \cos(x,y) \rightarrow L, A2 = \cos(x,y) \rightarrow L \\ A3 &= \cos(x,y) \rightarrow L, A4 = \cos(x,y) \rightarrow L \end{aligned} \quad (20)$$

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 14 shows the results of rotational angular joint features over the dance class.

After the completion of the contextual features portion, we concatenate all the sub-feature vectors into the main feature vector while (Algorithm 2) shows the detailed overview of the contextual features extraction approach.

Algorithm 2: Contextual Features Extraction

Input: HS: Human silhouette from RGB video data

Output: Contextual feature vectors($Cf_1, Cf_2, Cf_3, \dots, Cf_n$)

% feature vector for %

Contextual_features_vec $\leftarrow []$

CF_vecsize \leftarrow GetFeaturesVectorsize ()

% loop over human silhouettes %

For i = 1:K

Contextual_features_vec _interactions \leftarrow 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 (Cf1, Cf2, Cf3, ..., Cfn)

```

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.

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

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.

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

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 mpaii-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 15 shows the most accurate features results over the mp11-video pose, the COCO, and the 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.

$$TR_p = \sum_q w_{i.p.q} \times a_p + b_q \quad (23)$$

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

$$\sigma(TR_p) = \frac{e^{TR_p}}{\sum_{n=1}^n e^{TR_p}}, q = 1, \dots, n \quad (24)$$

Figure 16 shows the detailed process of CNN parameters learning over gait event detection and classification:



RESULTS

Dataset Descriptions

The Mp11-video pose data set is a large-scale dataset, which contains human activities and posture information-based videos. 21 different activities such as home activities, lawn, garden, sports, washing windows, picking fruit, and rock climbing. All the videos All the videos selected for our dataset collection have been recommended as YouTube top 10 videos in each activity. Figure 17 shows some example images of the Mp11-video-pose dataset.

The COCO (Common Objects in Context) dataset is based on multi-person tracking and object detection dataset, different activities contains in the COCO dataset They include Bicycling, Conditioning exercise, Dancing, Fishing and hunting, Music playing, Religious activities, Sports,

Transportation, Walking, Water activities, winter activities. Figure 18 shows some example images of the COCO dataset:

Pose track dataset is based on two main tasks including multi-person human pose estimation and analysis over a single frame. Videos and articulated tracking have been based on human posture estimation. Dataset mostly consists of complex videos such as crowded or team sports. Various activities have been covered in the pose track dataset. Figure 19 shows some example images of the pose track dataset.

Experiment I: The Landmarks Detection Accuracies

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

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

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

$$De = \frac{100}{n} \left[\sum_{n=1}^i \begin{cases} 1 & \text{if } De \leq 16 \\ 0 & \text{if } De > 16 \end{cases} \right] \quad (26)$$

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



(Table 3) represents the results of multi-person human body parts for the mpII-video-pose dataset. For identified body parts, we indicate with ✓ and for unidentified we adopted we use✖. We attained a detection accuracy for human1 -63.63%, human2 - 72.72%, human3 – 63.63%, human4- 72.72%, human5 - 72.72% and the mean detection accuracy of 69.09%.

(Table 4) represents the results of multi-person human body parts for the COCO dataset. For identified body parts, we indicated ✓ and for unidentified body parts we adopted ✖. We a attained detection accuracy for human1 - 81.81%, human2 - 72.72%, human3 - 72.72%, human4- 72.72%, human5 - 72.72%, and the mean detection accuracy of 74.54%.

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

Experiment II: Event Classification Accuracies

For gait event classification, we used a CNN-based deep learning approach. The design method is evaluated by the Leave One Subject Out (LOSO) cross-validation method. In Fig. 20, the results over the mpII-video-pose dataset show 90.90% gait event classification and detection accuracy. After this, we applied the deep belief network over the Olympic sports dataset and found the stochastic remote sensing event classification results. Figure 21 represents the confusion matrix for the COCO dataset with 89.09% mean accuracy for gait event classification. Finally, CNN is applied over the pose track dataset, with the mean gait event classification accuracy of 88.18%. Figure 22 represents the results in the shape of the confusion matrix for the pose track dataset.

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.

Experiment III: Comparison with Other Classification Algorithms

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.

Experiment IV: Comparison of our Proposed System with State-of-the-Art Techniques

Fan et al. (2015) developed a unique approach to estimate the human pose which is based on deep learning-based dual-source CNN. As the input they used patches of a given image and human body patches. After that they combined both contextual and local index values to estimate human posture with a better accuracy rate. Pishchulin et al. (2016) proposed a robust formulation as a challenge of subsection partitioning and labeling (SPLP). The SPLP structure, unlike previous two-stage methods that separated the identification and pose estimation measures, suggests the number of persons, certain poses, spatial proximity, including component level occlusions all at the same time. In Wei et al. (2016), convolutional position devices have an edge infrastructure for solving formal classifications based on computer vision that does not require visual type reasoning. We demonstrated that by transmitting increasingly refined confusion beliefs among points, a sequential framework consisting of convolutional networks is incapable of effectively training a structural component for the position. Jin et al. (2019) proposed SpatialNet and TemporalNet combined to form a single pose prediction and monitoring conceptual model: Body part identification and part-level temporal classification are handled by SpatialNet, while the contextual classification of human events is handled by TemporalNet. Bao et al. (2020) suggest a hand gesture identification-by-tracking system that

incorporates pose input into both the video human identification and human connection levels. A person's position prediction with pose descriptive statistics is used in the first level to reduce the impact of distracting and incomplete human identification in images. [Umer et al. \(2020\)](#) present a method for detecting people in video which depends on key feature connections. Rather than training the system to estimate key-point communications on video sequences, the system is equipped to estimate human pose utilizing personality on massive scale datasets. [Sun et al. \(2018\)](#) suggested a technique for features extracted in which they remove guided optical flow and use a CNN-based paradigm to identify and classify human events. [Rachmadi, Uchimura & Koutaki \(2016\)](#) described a method for dealing with event recognition and prevention using CNN and NNA (Network in Network Architecture) frameworks, which are the foundation of modern CNN. CNN's streamlined infrastructure, median, average, and commodity features are used to define human activities. [Zhu et al. \(2019\)](#) provide a detailed method for identifying incidents in security video. Throughout the TRECVID-SED 2016 test, their method outperformed others by a substantial margin by combining path modeling with deep learning. [\(Table 9\)](#) shows the gait event mean accuracy comparison with the other methods over the MPII, COCO and Pose track datasets.

CONCLUSION

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

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). Also, this work was supported by the Korea Medical Device Development Fund grant funded by the Korea

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

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Yazeed Yasin Ghadi conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Israr Akhter examined the experimental parameters, performed logical testing, analyzed the data as well as computational work, prepared figures and/or tables, and approved the final draft.
- Munkhjargal Gochoo analyzed the data and approved the final draft.
- Ahmad Jalal conceived and designed the experiments, figures and tables, authored or reviewed drafts of the paper, and approved the final draft.
- Kibum Kim conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability.
Code and data are available in the Supplemental Files.

Supplemental Information

Supplemental information for this article can be found online at

REFERENCES

- Aggarwal JK, Cai Q. 1999. Human Motion Analysis: A Review. *Computer Vision and Image Understanding*. DOI: [10.1006/cviu.1998.0744](https://doi.org/10.1006/cviu.1998.0744).
- Akhter I. 2020. 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.
- Akhter I, Jalal A, Kim K. Pose Estimation and Detection for Event Recognition using Sense-Aware Features and Adaboost Classifier.
- Akhter I, Jalal A, Kim K. 2021. Adaptive Pose Estimation for Gait Event Detection Using Context-Aware Model and Hierarchical Optimization. *Journal of Electrical Engineering & Technology*:1–9.
- Amft O, Tröster G. 2008. Recognition of dietary activity events using on-body sensors. *Artificial Intelligence in Medicine*. DOI: [10.1016/j.artmed.2007.11.007](https://doi.org/10.1016/j.artmed.2007.11.007).
- Bao Q, Liu W, Cheng Y, Zhou B, Mei T. 2020. Pose-guided tracking-by-detection: Robust multi-person pose tracking. *IEEE Transactions on Multimedia* 23:161–175.
- Einfalt M, Dampeyrou C, Zecha D, Lienhart R. 2019. 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*. New York, New York, USA: ACM Press, 42–50. DOI: [10.1145/3321111.3321112](https://doi.org/10.1145/3321111.3321112).

- 10.1145/3347318.3355525.
- Fan X, Zheng K, Lin Y, Wang S. 2015.** Combining local appearance and holistic view: Dual-source deep neural networks for human pose estimation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1347–1355.
- Feng Q, Gao C, Wang L, Zhao Y, Song T, Li Q. 2020.** Spatio-temporal fall event detection in complex scenes using attention guided LSTM. *Pattern Recognition Letters*. DOI: 10.1016/j.patrec.2018.08.031.
- Franco A, Magnani A, Maio D. 2020.** A multimodal approach for human activity recognition based on skeleton and RGB data. *Pattern Recognition Letters*. DOI: 10.1016/j.patrec.2020.01.010.
- Franklin RJ, Mohana, Dabbagol V. 2020.** Anomaly Detection in Videos for Video Surveillance Applications using Neural Networks. In: *Proceedings of the 4th International Conference on Inventive Systems and Control, ICISC 2020*. DOI: 10.1109/ICISC47916.2020.9171212.
- Gochoo M, Akhter I, Jalal A, Kim K. 2021.** Stochastic Remote Sensing Event Classification over Adaptive Posture Estimation via Multifused Data and Deep Belief Network. *Remote Sensing* 13. DOI: 10.3390/rs13050912.
- Golestani N, Moghaddam M. 2020.** Human activity recognition using magnetic induction-based motion signals and deep recurrent neural networks. *Nature Communications*. DOI: 10.1038/s41467-020-15086-2.
- Hassner T, Basri R. 2006.** Example based 3D reconstruction from single 2D images. In: *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*. 15.
- Hu Y, Jiang D, Yan S, Zhang L, others. 2004.** Automatic 3D reconstruction for face recognition. In: *Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings*. 843–848.
- Jalal A, Akhtar I, Kim K. 2020.** Human Posture Estimation and Sustainable Events Classification via Pseudo-2D Stick Model and K-ary Tree Hashing. *Sustainability* 12:9814.
- Jalal A, Khalid N, Kim K. 2020.** Automatic recognition of human interaction via hybrid descriptors and maximum entropy markov model using depth sensors. *Entropy*. DOI: 10.3390/E22080817.
- Jiang YG, Dai Q, Mei T, Rui Y, Chang SF. 2015.** Super Fast Event Recognition in Internet Videos. *IEEE Transactions on Multimedia*. DOI: 10.1109/TMM.2015.2436813.
- Jin S, Liu W, Ouyang W, Qian C. 2019.** Pose-guided tracking-by-detection: Robust multi-person pose tracking In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Khalid N, Gochoo M, Jalal A, Kim K. 2021.** Modeling Two-Person Segmentation and Locomotion for Stereoscopic Action Identification: A Sustainable Video Surveillance System. *Sustainability* 13:970.
- Khan MA, Javed K, Khan SA, Saba T, Habib U, Khan JA, Abbasi AA. 2020.** Human action recognition using fusion of multiview and deep features: an application to video surveillance. *Multimedia Tools and Applications*. DOI: 10.1007/s11042-020-08806-9.
- Khan MH, Zöller M, Farid MS, Grzegorzec M. 2020.** Marker-based movement analysis of human body parts in therapeutic procedure. *Sensors (Switzerland)*. DOI: 10.3390/s20113312.
- van der Kruk E, Reijne MM. 2018.** Accuracy of human motion capture systems for sport applications; state-of-the-art review. *European Journal of Sport Science*. DOI:

[10.1080/17461391.2018.1463397](https://doi.org/10.1080/17461391.2018.1463397).

- Li A, Miao Z, Cen Y, Zhang X-P, Zhang L, Chen S. 2020.** Abnormal event detection in surveillance videos based on low-rank and compact coefficient dictionary learning. *Pattern Recognition* 108:107355.
- Liu J, Luo J, Shah M. 2009.** Recognizing realistic actions from videos in the Wild. In: *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009*. DOI: [10.1109/CVPRW.2009.5206744](https://doi.org/10.1109/CVPRW.2009.5206744).
- Lohithashva BH, Aradhya VNM, Guru DS. 2020.** Violent video event detection based on integrated LBP and GLCM texture features. *Revue d'Intelligence Artificielle* 34:179–187.
- Mokhlespour Esfahani MI, Zobeiri O, Moshiri B, Narimani R, Mehravar M, Rashedi E, Parnianpour M. 2017.** Trunk motion system (TMS) using printed body worn sensor (BWS) via data fusion approach. *Sensors (Switzerland)*. DOI: [10.3390/s17010112](https://doi.org/10.3390/s17010112).
- Pishchulin L, Insafutdinov E, Tang S, Andres B, Andriluka M, Gehler P V, Schiele B. 2016.** Deepcut: Joint subset partition and labeling for multi person pose estimation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4929–4937.
- Rachmadi RF, Uchimura K, Koutaki G. 2016.** Combined convolutional neural network for event recognition. In: *Proceedings of the Korea-Japan Joint Workshop on Frontiers of Computer Vision*. 85–90.
- Sun S, Kuang Z, Sheng L, Ouyang W, Zhang W. 2018.** Optical Flow Guided Feature: A Fast and Robust Motion Representation for Video Action Recognition. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. DOI: [10.1109/CVPR.2018.00151](https://doi.org/10.1109/CVPR.2018.00151).
- Tahir SB. 2020.** A Triaxial Inertial Devices for Stochastic Life-Log Monitoring via Augmented-Signal and a Hierarchical Recognizer. Ph.D. Thesis, Air University, Islamabad, Pakistan, December 2020.
- Tahir SB, Jalal A, Kim K.** IMU Sensor based Automatic-Features Descriptor for Healthcare Patient's daily life-log Recognition.
- Uddin MZ, Thang ND, Kim JT, Kim T-S. 2011.** Human activity recognition using body joint-angle features and hidden Markov model. *Etri Journal* 33:569–579.
- Ullah A, Muhammad K, Haq IU, Baik SW. 2019.** Action recognition using optimized deep autoencoder and CNN for surveillance data streams of non-stationary environments. *Future Generation Computer Systems*. DOI: [10.1016/j.future.2019.01.029](https://doi.org/10.1016/j.future.2019.01.029).
- Umer R, Doering A, Leibe B, Gall J. 2020.** Self-supervised keypoint correspondences for multi-person pose estimation and tracking in videos. *arXiv preprint arXiv:2004.12652*.
- Rehman MA, Raza H, Akhter I. 2018.** 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*. Acavent,. DOI: [10.33422/4kiconf.2018.12.24](https://doi.org/10.33422/4kiconf.2018.12.24).
- Wang Y, Du B, Shen Y, Wu K, Zhao G, Sun J, Wen H. 2019.** EV-gait: Event-based robust gait recognition using dynamic vision sensors. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. DOI: [10.1109/CVPR.2019.00652](https://doi.org/10.1109/CVPR.2019.00652).
- Wang Y, Mori G. 2008.** Multiple tree models for occlusion and spatial constraints in human pose estimation. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. DOI: [10.1007/978-3-540-](https://doi.org/10.1007/978-3-540-)

88690-7-53.

Wei S-E, Ramakrishna V, Kanade T, Sheikh Y. 2016. Multi-Person Articulated Tracking With Spatial and Temporal Embeddings, Proceedings of the IEEE/CVF Conference on Computer Vision . In: *2016 IEEE Conference on*.

Yu J, Lei A, Hu Y. 2019. Soccer video event detection based on deep learning. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. DOI: [10.1007/978-3-030-05716-9_31](https://doi.org/10.1007/978-3-030-05716-9_31).

Zheng Y, Gu S, Edelsbrunner H, Tomasi C, Benfey P. 2011. Detailed reconstruction of 3D plant root shape. In: *2011 International Conference on Computer Vision*. 2026–2033.

Zhu Y, Zhou K, Wang M, Zhao Y, Zhao Z. 2019. A comprehensive solution for detecting events in complex surveillance videos. *Multimedia Tools and Applications*. DOI: [10.1007/s11042-018-6163-6](https://doi.org/10.1007/s11042-018-6163-6).

Zou Y, Shi Y, Shi D, Wang Y, Liang Y, Tian Y. 2020. Adaptation-Oriented Feature Projection for One-shot Action Recognition. *IEEE Transactions on Multimedia*. DOI: [10.1109/tmm.2020.2972128](https://doi.org/10.1109/tmm.2020.2972128).

Figure 1

The proposed system model's structural design.

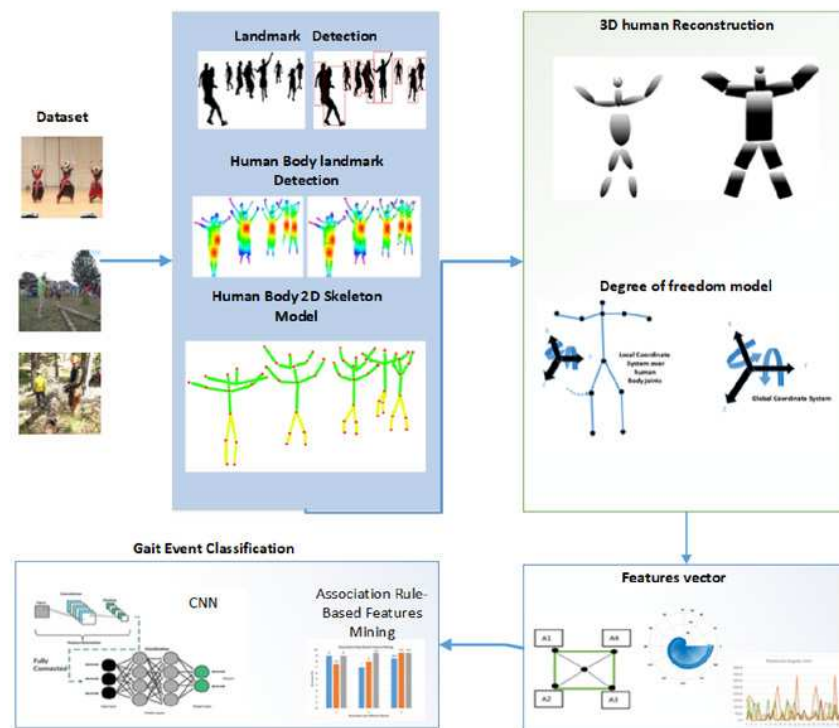


Figure 1. The proposed system model's structural design.

Figure 2

Results of different background subtraction techniques along with the original image

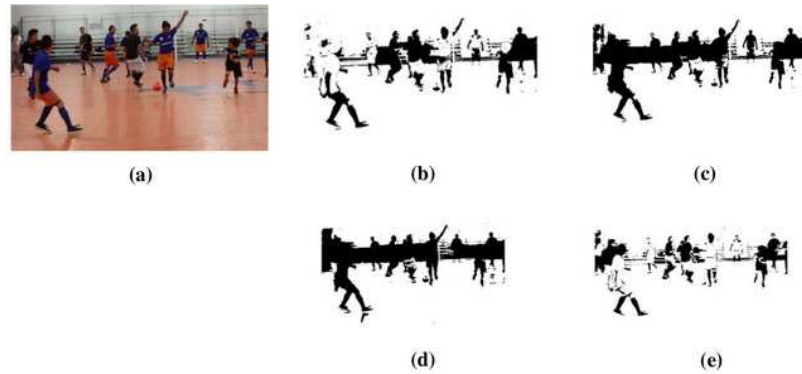


Figure 2. Results of different background subtraction techniques along with the original image. (a) Original image (b) change detection (c) floor detection (d) Markov random field and (e) spatial-temporal differencing.

Figure 3

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

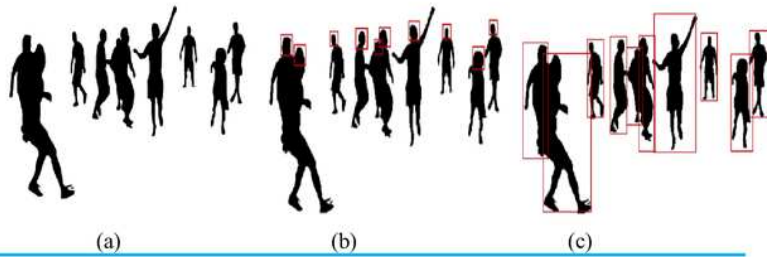


Figure 3. Results of (a) Optimized human silhouette (b) human head detection (c) human detection in RGB videos and image sequences.

Figure 4

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

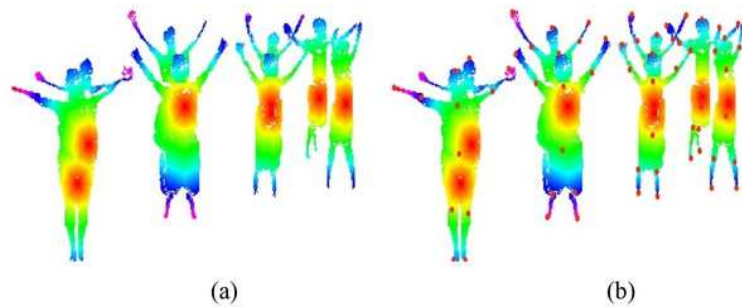


Figure 4. Human body landmark detection results (a) presents the landmark results using an HSV color map, (b) presents the eleven human body points.

Figure 5

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

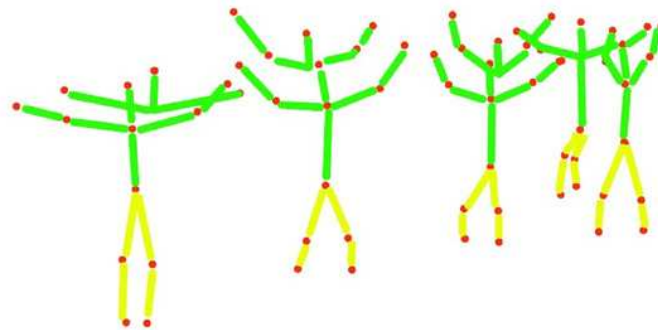


Figure 5. The Human 2D skeleton model results in over eleven human body parts.

Figure 6

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

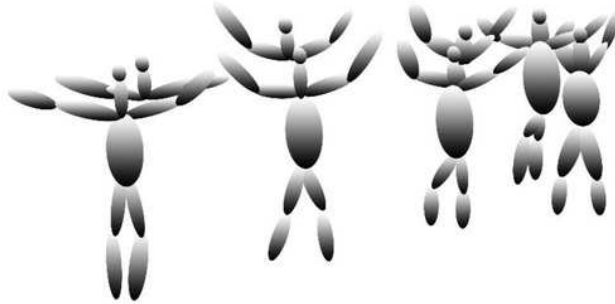


Figure 6. The results of the computational model with ellipsoids over human body points.

Figure 7

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.

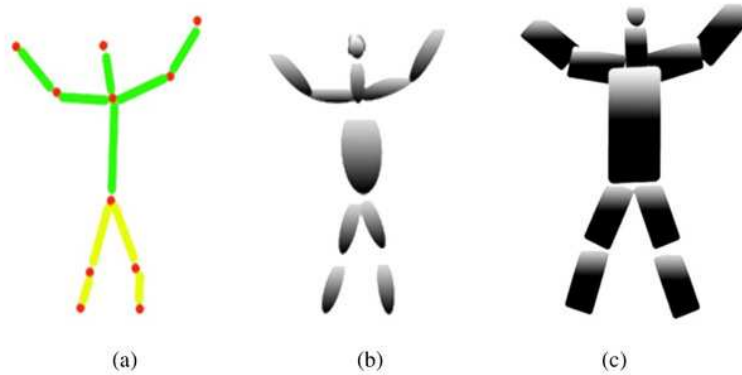


Figure 7. The results of the synthetic model with super quadrics over human body points. (a) Human 2D skeleton, (b) computational model with ellipsoids (c) synthetic model with super quadrics.

Figure 8

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.

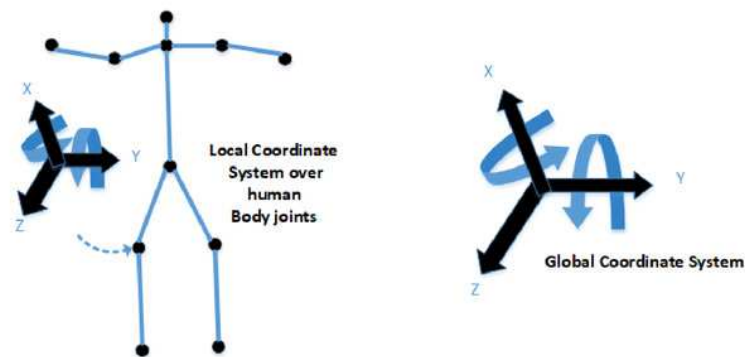


Figure 8. The theme concept of local and global coordinate systems. The left side shows the local coordinate system over the human left knee; the right side shows the DOF based global coordinate system.

Figure 9

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

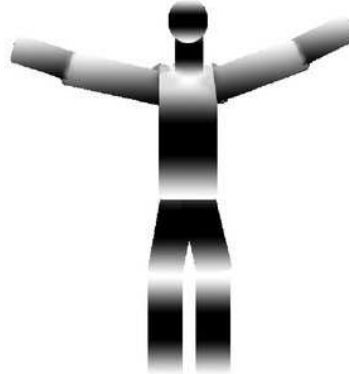


Figure 9. The results of the 3D ellipsoid reconstruction over the synthetic model with super quadrics and joint angle estimation.

Figure 10

Few DOF results examples.

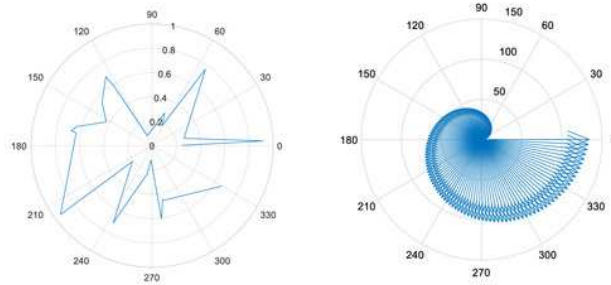


Figure 10. Few DOF results examples.

Figure 11

Periodic motion results.



Figure 11. Periodic motion results.

Figure 12

Results of non-periodic motion.



Figure 12. Results of non-periodic motion.

Figure 13

The results of motion direction flow over the basketball video.



Figure 13. The results of motion direction flow over the basketball video.

Figure 14

Rotational angular joint results and the pattern of rotational angels.

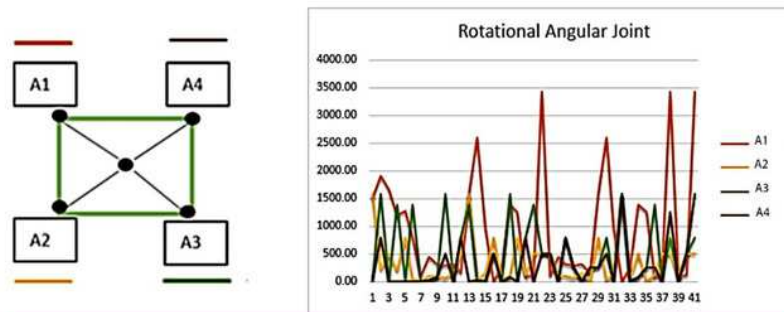


Figure 14. Rotational angular joint results and the pattern of rotational angels.

Figure 15

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

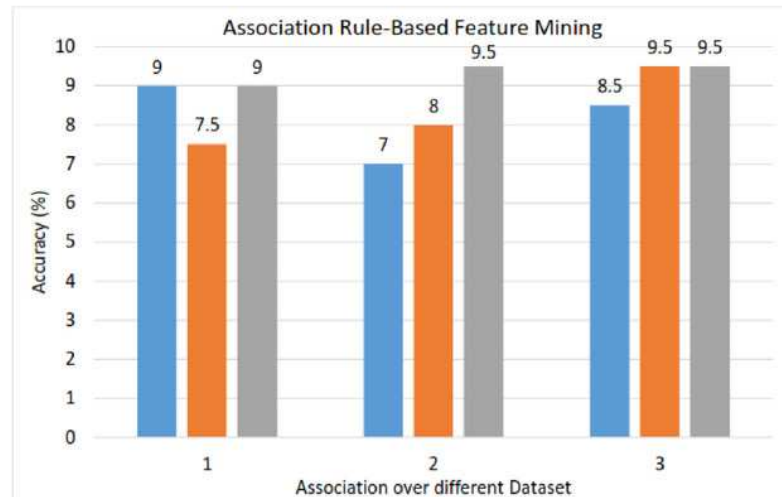


Figure 15. The most accurate features results via the rule-based features mining approach over the mpii-video pose, COCO, and Pose track datasets.

Figure 16

CNN model overview.

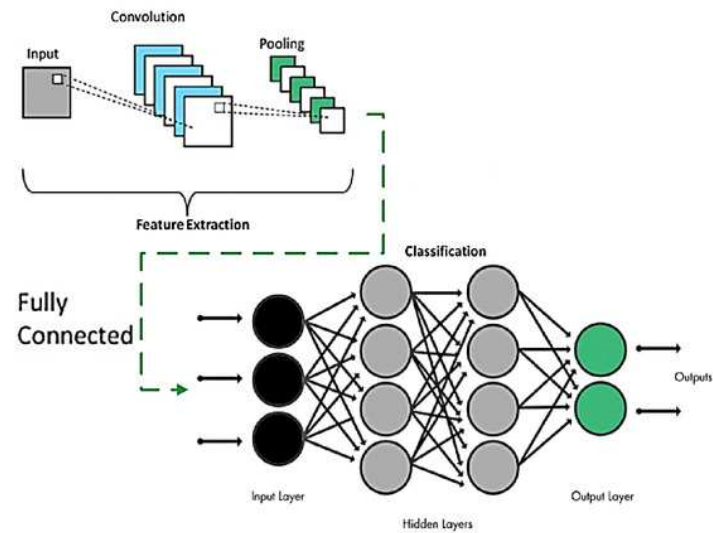


Figure 16. CNN model overview.

Figure 17

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



Figure 17. A few example images of the mpii-video-pose dataset.

Figure 18

A few example images from the COCO dataset.



Figure 18. A few example images from the COCO dataset.

Figure 19

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

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.

Figure 20

A few example images of the postmark dataset.



Figure 19. A few example images of the postmark dataset.

Figure 21

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 22. Confusion matrix results using CNN over the Pose track dataset

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.

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 21. Confusion matrix results using CNN over the COCO dataset.

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.

Box 1(on next page)

Human Silhouette Optimization

(Algorithm 1) shows the detailed procedure of silhouette optimization.

Algorithm 1: 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 $WP_{pixel\ information\ of\ frame\ 1} = WP_{information\ of\ frame\ 2}$

$WP_{pixel\ information\ of\ frame\ 3} = WP_{pixel\ information\ of\ frame\ 1}$

End

If WP is inadequate for all inputs

If pixel information is equal with SP

$OS = WP_{pixel}$

End

End

End

Table 1 (on next page)

Comprehensive review of relevant research

Table 1 includes a comprehensive review of recent research in this area.

Table 1 Comprehensive review of relevant research.

Human 2D posture analysis and event detection

Methods	Main contributions
<i>Liu, Luo & Shah (2009)</i>	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.
<i>Khan et al. (2020)</i>	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.
<i>Zou et al. (2020)</i>	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.
<i>Franco, Magnani & Maio (2020)</i>	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.
<i>Ullah et al. (2019)</i>	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.
<i>van der Kruk & Reijne (2018)</i>	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.
<i>Wang & Mori (2008)</i>	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.
<i>Amft & Tröster (2008)</i>	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.
<i>Wang et al. (2019)</i>	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.

<i>Jiang et al.</i> (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.
<i>Li et al.</i> (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.
<i>Einfalt et al.</i> (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.
<i>Yu, Lei & Hu</i> (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.
<i>Franklin, Mohana & Dabbagol</i> (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.
3D human posture analysis and event detection	
<i>Aggarwal & Cai</i> (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.
<i>Hassner & Basri</i> (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.
<i>Hu et al.</i> (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.
<i>Zheng et al.</i> (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.
<i>Uddin et al. (2011)</i>	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).
<i>Lohithashva, Aradhya & Guru (2020)</i>	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.
<i>Feng et al. (2020)</i>	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.
<i>Khan et al. (2020)</i>	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.
<i>Mokhlespour Esfahani et al. (2017)</i>	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.
<i>Golestani & Moghaddam, (2020)</i>	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.

Table 2 (on next page)

Human body parts recognition and detection accuracy

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

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

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

Table 3(on next page)

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

(Table 3) represents the results of multi-person human body parts for the mpII-video-pose dataset. For identified body parts, we indicate with ✓ and for unidentified we adopted we use✕. We attained a detection accuracy for human1 -63.63%, human2 - 72.72%, human3 – 63.63%, human4- 72.72%, human5 - 72.72% and the mean detection accuracy of 69.09%.

Table 3: 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%	72.72%	63.63%	72.72%	72.72%
Mean accuracy = 69.09%					

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

Table 4(on next page)

Human body parts results of multi-person for COCO dataset

(Table 4) represents the results of multi-person human body parts for the COCO dataset. For identified body parts, we indicated ✓ and for unidentified body parts we adopted ✕. We attained detection accuracy for human1 - 81.81%, human2 - 72.72%, human3 - 72.72%, human4- 72.72%, human5 - 72.72%, and the mean detection accuracy of 74.54%.

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%					

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

Table 5(on next page)

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

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

Table 5: 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%
Mean accuracy = 65.45%					

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

Table 6 (on next page)

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

(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.

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	Precisi on	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

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

Table 7 (on next page)

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		
	Precisi on	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

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.

Box 2(on next page)

Contextual Features Extraction

(Algorithm 2) shows the detailed overview of the contextual features extraction approach.

Algorithm 2: Contextual Features Extraction

Input: HS: Human silhouette from RGB video data

Output: Contextual feature vectors($Cf_1, Cf_2, Cf_3, \dots, Cf_n$)

% feature vector for %

Contextual_features_vec \leftarrow []

CF_vecsize \leftarrow GetFeaturesVectorsize ()

% loop over human silhouettes %

For i = 1:K

Contextual_features_vec _ interactions \leftarrow Get _ Contextual_features_vec (interactions)

% extracting DOF, periodic motion, non periodic motion, motion direction and flow,

Rotational angular joint%

DOF \leftarrow ExtractDOF(Contextual_features_vec _ interactions)

PeriodicMotion \leftarrow ExtractPeriodicMotion (Contextual_features_vec _ interactions)

NonPeriodicMotion \leftarrow ExtractNonPeriodicMotion (Contextual_features_vec _ interactions)

MotionDirectionandFlow \leftarrow ExtractMotionDirectionandFlow(Contextual_features_vec_interactions)

RotationalAngularJoint \leftarrow ExtractRotationalAngularJoint(Contextual_features_vec _ interactions)

Contextualvectors \leftarrow GetCFeaturevector

FVectors.append (CF_vectors)

End

Contextualvectors \leftarrow Normalize (Contextual_features_vec)

return Contextual_features_vec ($Cf_1, Cf_2, Cf_3, \dots, Cf_n$)

Table 8(on next page)

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	Precisi on	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

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.

Table 9(on next page)

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

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

Table 9. Gait event mean accuracy comparison with the other methods over the MPII, COCO and Pose track datasets.

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