Tech Science Press

Self-Care Assessment for Daily Living Using Machine Learning Mechanism

Mouazma Batool[1], Yazeed Yasin Ghadi[2], Suliman A. Alsuhibany[3], Tamara al Shloul[4], Ahmad Jalal[1] and Jeongmin Park[5,*]

[1]Department of Computer Science, Air University, Islamabad, 44000, Pakistan
[2]Department of Computer Science and Software Engineering, Al Ain University, Al Ain, 15551, UAE
[3]Department of Computer Science, College of Computer, Qassim University, Buraydah, 51452, Saudi Arabia
[4]Department of Humanities and Social Science, Al Ain University, Al Ain, 15551, UAE
[5]Department of Computer Engineering, Korea Polytechnic University, 237 Sangidaehak-ro Siheung-si, Gyeonggi-do, 15073, Korea
*Corresponding Author: Jeongmin Park. Email: jmpark@kpu.ac.kr
Received: 12 November 2021; Accepted: 11 January 2022

Abstract: Nowadays, activities of daily living (ADL) recognition system has been considered an important field of computer vision. Wearable and optical sensors are widely used to assess the daily living activities in healthy people and people with certain disorders. Although conventional ADL utilizes RGB optical sensors but an RGB-D camera with features of identifying depth (distance information) and visual cues has greatly enhanced the performance of activity recognition. In this paper, an RGB-D-based ADL recognition system has been presented. Initially, human silhouette has been extracted from the noisy background of RGB and depth images to track human movement in a scene. Based on these silhouettes, full body features and point based features have been extracted which are further optimized with probability based incremental learning (PBIL) algorithm. Finally, random forest classifier has been used to classify activities into different categories. The n-fold cross-validation scheme has been used to measure the viability of the proposed model on the RGBD-AC benchmark dataset and has achieved an accuracy of 92.71% over other state-of-the-art methodologies.

Keywords: Angular geometric features; decision tree classifier; human activity recognition; probability based incremental learning; ridge detection

1 Introduction

In the world of artificial intelligence, Activities of Daily Living (ADL) has gained much of research interest for assisted living environments. The fundamental activities of daily life (ADL) are related to personal self-care and are defined as the basic activities of daily life. Doctor Deaver and Brown were the first to establish the concept of ADL in America [1]. The ADL is a medical instrument used to evaluate old persons, those with mental illnesses, and others. It is also gaining popularity in the measurement of human body motion [2]. Nowadays, videos with huge volumes have a lot of

irrelevant content, therefore accurate recognition is required. Various sensors have been utilized in ADLs, namely, wearable sensors, optical sensors, and Radio Frequency Identification (RFID) but optical sensors-based systems have gained a lot of interest. Most of the previous work on human life-logging activity has been based on Red Green Blue (RGB) video and RFID sensors [3]. The RGB video systems achieved a very low accuracy of 78.5% even in the absence of clutter. On the other hand, the RFID method is generally too intrusive as it needs RFID tags on the people [4]. The accurate and effective recognition of ADLs plays an important role in determining a person's condition in realistic environment.

Different techniques have been developed for ADLs in recent years [5]. The space-time feature in-combination with bag of features is the state-of-the-art method for video representation [6]. The feature-based methods have usually been used for object recognition, which is further successfully tested for activity recognition [7]. The sparse-based action recognition system reduces the overall computational cost but results in performance degradation [8].

In this article, a unique machine learning framework is adopted to recognize the ADL system using the sequence of RGB and depth images. First, we extracted human RGB silhouette by applying background subtraction over the unwanted areas of images, and the silhouette has been extracted from the depth images by using the morphological operation model. Later on, system is designed that consists of ridge and angular-geometric features for feature extraction. Finally, combination of probability-based incremental learning (PBIL) and random forest classifier has been used for feature optimization and classification. The RGBD-AC publicly available benchmark dataset has been used for the validation of our results against state-of-the-art models. The proposed technique is based on five different aspects, silhouette extraction from the images, ridge detection, angular-geometric features modeling of a human silhouette, optimization, and classification. In addition, the proposed model has been applied over RGBD-AC dataset and achieved substantial results over other state-of-the-art methodologies.

The rest of this paper has been structured as: Section 2 comprehends related work, Section 3 covers the architectural representation of an anticipated model. Section 4 depicts the performance evaluation of the proposed work, Section 5 contains related discussions. Section 6 provides the conclusion and future directions on the proposed methodology.

2 Related Work

For ages, RGB and depth images have been separately anticipated for ADL recognition systems. Lee et al. [9] has used RGB images for the classification of human actions. The OpenPose and 3D-baseline open source libraries has been used to extract skeleton joint features of skeleton. The CNN classifier has been finally applied on the extracted features. Their model has achieved an accuracy of 70% on NTU-RGBD dataset. In Mahmood et al. [10] WHITE STAG model is proposed for the RGB dataset. They have applied space-time and angular-geometric features over full-body silhouette and skeleton joint features. The performance of the model has been evaluated with a leave-one-out cross-validation scheme and achieved an overall accuracy of 87%. Varshney et al. [11] have utilized a skeleton tracking algorithm for the recognition of discrete skeleton positions on a real time RGB-D data streams. The proposed methodology has tested on 14 different activities that has been performed by two person in home environment and lecture hall. The main limitation in this model is that the proposed model has only tested on limited number of subject and has not been

tested on any benchmark dataset. Si et al. [12] have extracted spatial and temporal features of RGB skeleton sequences images and CNN has applied to classify the discriminative features for human activity recognition. This model has tested over 3 benchmark datasets and has achieved an accuracy of 85% over SYSU-3D datasets. Kim et al. [13] have targeted human pose estimation problem using the improved architecture of hourglass network. The conventional hourglass network requires high computational power which is reduced by several skip connections. Hence the architecture of the network is improved with minimal modification. The proposed model has achieved an accuracy of 90.8%. Madhuranga et al. [14] employed CNN in-depth images that were previously used for RGB video for silhouette segmentation. They have developed the model for real-time recognition, which is capable of recognizing ADL on commodity computers. Khalid et al. [15] proposed K-ary Tree Hashing classifier to categorize NTU RGB-D sequential images. Buys et al. [16] extracted human skeleton from complex scenes using RGB-D sensors. The model was designed for color and depth image sequence structure in an environment. The results are further fed to RDF (Random Decision Forest) for cluster-based learning. Fu et al. [17] presented a Bayesian conditional probability problem over RGB-D based dataset. They have tested their model on twelve different activities performed by four different people and achieved good recognition accuracy.

The main drawback in RGB methodology is that depth information has been lost during 3D to the 2D project process. Moreover, the depth information depends only on the distance, which leads to difficulty in finding invariant features of objects. The RGB-D sensors can capture color and depth information simultaneously. The color and appearance information has been contained by RGB image whereas variations in color and illumination has been depicted from depth image. Thus, RGB-D images have gained popularity in the ADL recognition mechanism. In this paper, we proposed a novel mechanism for the RGB-D image-based ADL recognition system. In this work, RGB and depth image sequences has been adopted to bring robustness into our ADL model. For accurate pre-processing of data, separate methods have been taken into account for RGB and depth images. While, full-body features and angle-based geometric features bring more strength to the model. Moreover, PBIL and random forest classifier enhanced the classification of activities. The major contribution in this work are: (1) pre-processing of RGB and depth images effectively, (2) integration of full-body feature and angle based geometric feature for the feature extraction process, (3) using PBIL and decision classifier together for accurate classification of action recognition, and (4) validation of proposed model over RGBD-AC dataset.

3 Material and Methods

The proposed methodology has taken RGB and depth images together as input to the system. In the pre-processing step, the human silhouette has been extracted from the raw images (RGB and depth) by applying background subtraction on RGB images, and morphological operation on depth images. Further, full-body features and angular-geometric features have been extracted from the silhouette. Ridge has been used to detect full-body features and the angular-geometric feature has been applied on the skeleton model to extract point-based features of human silhouette. Moreover, the probability-based incremental learning algorithm has been applied in order to symbolize the full-body features and angular-geometric features. Lastly, a decision tree classifier has been applied to the resultant optimized features. The overview of the model has been depicted in Fig. 1.

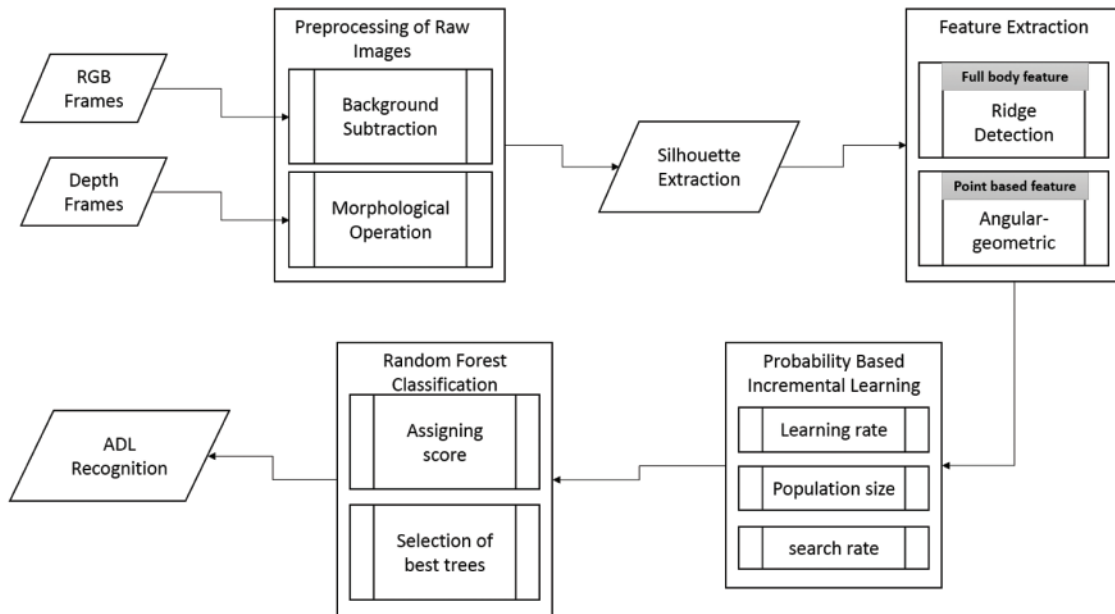**Figure 1:** The technical architecture diagram for the proposed ADL model

3.1 Preprocessing of Raw Images

First, all the raw images (RGB and depth) have been processed through the image normalization technique in order to enhance the quality of RGB-D images. Then, image contrast has been adjusted and a histogram equation has been applied to the image to uniformly distribute the intensity values through the entire image. The median filter has been applied to efficiently remove noise from the image [18]. It works by replacing the pixels with neighbouring pixels [19]. Finally, the region of interest (ROI) has been extracted from the image that have been segmented from their background, which has been described as below.

3.1.1 Background Subtraction for RGB Images

The silhouette of RGB image has been extracted through the background subtraction procedure [20]. The frame difference has been applied to the images from which current frame has been subtracted from the previous frames in each iteration [21]. The pixels of current frame $P[I(t)]$ have been subtracted from the pixels of background images $P[I(t-1)]$ at the time t that can be depicted using Eq. (1) as:

$$P[T] = P[I(t)] - P[I(t-1)] \tag{1}$$

where $P[T]$ is the resultant frame of background subtraction. The output of the frame contains a human silhouette. The resultant human silhouette has been further processed for foreground extraction by specifying a threshold Th as described in Eq. (2).

$$P[I(t)] - P[I(t+1)] > Th \tag{2}$$

The threshold of the images has been selected by using Otsu's threshold technique [22]. In Otsu's method, the subtracted human silhouette frame has been first converted to a grayscale image, and then Otsu automatically selects the best value for threshold Th. The best value is the value that efficiently differentiates the white foreground pixels from black background pixels. The Otsu's threshold has been

iteratively applied to the images until the best threshold value is obtained [23]. The threshold *Th* has been then further utilized to convert grayscale image to the binary image. The resultant images have been depicted in Fig. 2 as follows.

Figure 2: Human silhouette detection of RGB images: (a) Original image, (b) Subtracted frame, (c) Binary silhouette, and (d) Resultant RGB silhouette

3.1.2 Morphological Operations for Depth Images

The human silhouette from the depth images have been obtained through a morphological operation [24]. Firstly, binary images have been obtained from the depth images by applying threshold-based segmentation. The morphological operation has been applied to the segmented images by using the binary dilation process. The binary dilation process works by adding pixels to human edges and binary erosion has been applied on the resultant images that removes the boundary pixels [25]. The description for binary dilation and erosion has been portrayed in Eqs. (3)–(4) as follows:

$$Y \oplus Z = \{p | (\hat{Z})_z \cap Y \neq \emptyset\} \tag{3}$$

$$Y - Z = \{p | (Z) \subseteq Y\} \tag{4}$$

where p is the location of pixels for element Z. The reflection of element Z is $\hat{Z}$ that joins with foreground pixels denoted by Y during translation to Z. Through this procedure, only the main object within the frame has been maintained. Later on, Canny edge detection has been applied on the resultant frame in order to separate foreground pixels from the background pixels. Finally, a human silhouette has been obtained by removing smaller area objects from the binary image. The morphological results are shown in Fig. 3.
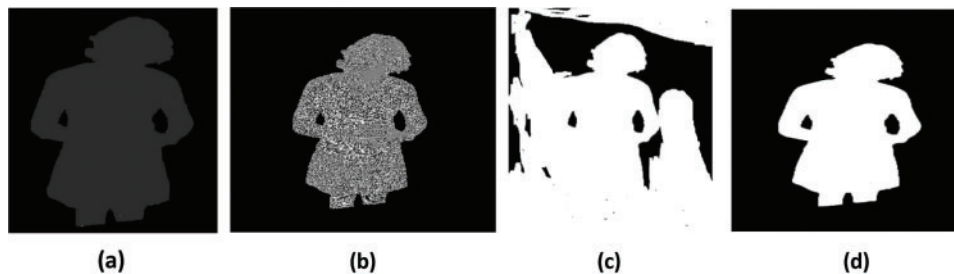

Figure 3: Human silhouette detection of depth images by applying morphological operation: (a) Erosion and dilation results for an image, (b) Edge detection, (c) Binary silhouette, and (d) Depth silhouette

3.2 Ridge Detection Features

The ridge detection of human silhouette comprises of two steps that are binary edge extraction and ridge data generation [26]. In the binary edge extraction step, the binary edges have been extracted from the RGB and depth silhouette obtained in the described pre-processing stage. The distance maps have been produced on the edges by using distance transform. While, in ridge data generation step, the local maximal has been obtained from the pre-computed maps, which produced ridge data within the binary edges [27]. A further description of binary edge detection and generation of ridge data is described below:

3.2.1 Edge Detection of Binary Image

The edge of the binary image has been calculated by computing the binary edge information of human silhouette. The binary edge information of RGB and depth silhouette is obtained using a window searching mechanism that computes statistical values of intensities from their nearest neighbor's pixel values [28]. As a result, correct edge connectivity of bounded body structure has been obtained using Eq. (5).

$$B(I) = \{P_c \in I | \exists P_i, |D(P_i) - D(P_c)| > \delta_e\}, P_i \in \{P_{c-1}, P_{c+1}, P_{c-w}, P_{c+w}\} \tag{5}$$

where P_c depicts the center pixel, and P_i represents the neighboring pixels. Both P_c and P_i have been compared in order to compute the intensity value of pixels. Later on, a distance map has been further obtained by applying the distance transformation technique on the resultant binary edges.

3.2.2 Ridge Data Generation

In this step, the distance maps of the binary edges have been calculated that have been utilized later to calculate the local maximal called ridge data [29]. The binary edges of the ridge data have been calculated using Eq. (6) as:

$$R(I) = \left\{ P_c \in I \frac{D_t(X_i)}{D_t(X_c)} \right\} < \delta_r \tag{6}$$

where D_t is the distance map that computes the center point of current pixel with their respective neighboring pixels. The ridge data eliminate noise around the edge data of (See Fig. 4) human silhouette and hence refined skeleton has been obtained which is represented in Fig. 5.

(a)   (b)

Figure 4: Ridge extraction of human silhouette (a) Binary edge extraction and (b) Ridge data generation
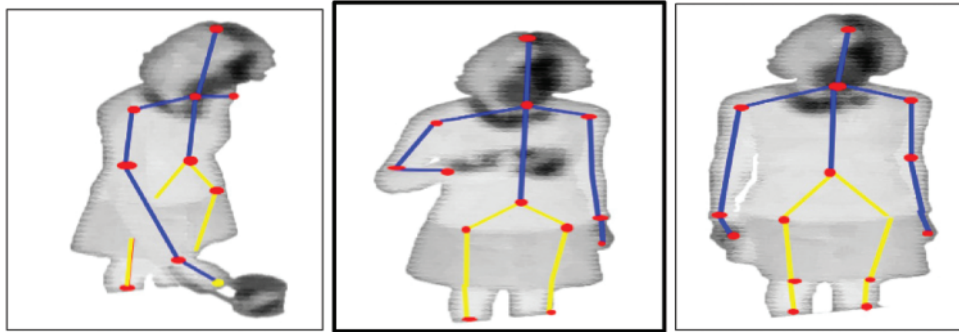
Figure 5: Set of skeletal points that detect the joint on human body skeleton

3.3 Skeleton Point Modeling

In skeleton point modeling, firstly, the contours of a human silhouette have been calculated that detect the outer pixels of human shape [30]. Secondly, torso has been traced by approximating the center point of calculated human contours [31], which is depicted in Eq. (7).

$$Ts = \frac{|H_x + H_y + H_z|}{\sqrt{H^2}} \tag{7}$$

where H_x and H_y are the x and y coordinates of the contours. While Ts is the torso of centroid points. To detect the height and width of human shape, pixel-wise search is applied to measure the size of head that is 1/6.5th of height of human silhouette, see Eq. (8).

$$HP_h^{fm} = HP_h^{fm-1} + \Delta HP_h^{fm-1} \tag{8}$$

where HP_h^{fm} is the position of head at a given frame f_m, which has been achieved by taking the difference of consecutive frame sequences. The position of the knee has been recognized using Eq. (9), which is located at the midpoint of the feet and torso joint.

$$HP_K^{fm} = (HP_F^{fm-1} + \Delta HP_{TC}^{fm-1})/2 \tag{9}$$

Finally, thirteen key points of e body has been detected (see Algorithm 1) to get the optimal location of body points.

Algorithm 1: Skeleton Point Modeling

Input: Extracted Human Silhouettes (HS)

Output: Thirteen body key points of human shape (hs): {left (Lf) and right (Rt) shoulders, hands, elbows, knees, feet} and {head (Hd), mid, neck (Nk), Height (Ht) and Width (Wd) of human shape (hs)}

 do

  HR ← Getvectors(HeadPoint(HS), UpperPoint(Hd), EndHeadPoint(Hd), Bottom(hs))
  Img_Mid ← Calculate mid(Ht, Wd)/2
  Img_Foot ← Calculate Bottom(hs) & search(Lf, Rt)
  Img_K ← Calculate mid(Img_Mid, Img_Foot)
  Img_H ← Hd & search(Lf, Rt)
  Img_S ← search(Hd, Nk) & search(Lf, Rt)

Algorithm 1: Continued

   Img_E ← *mid(Img_H, Img_S)*
 While (optimal locations of HS)
 return thirteen body key points

The skeletonization process detects thirteen body key points in three main body segments including lower body (*LB*), upper body (*UB*), and mid body (*MB*). *UB* connects the shoulders HP_{sh}, hands HP_{ha}, head HP_{hd}, elbows HP_{el}, and neck HP_{nk}. While, *MB* links mid-points of the body, which is denoted by HP_M. Finally, *LB* relates knees HP_{kn} and foot HP_{ft} as described in Eqs. (10)–(12). Each body key point perform specific function at specific time *t*.

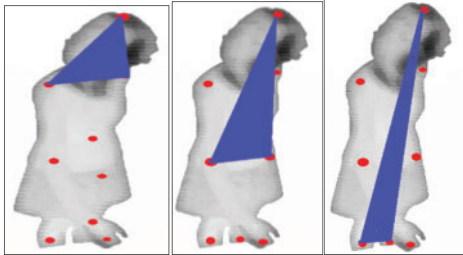$$UB_t = HP_{hd} + HP_{nk} + HP_{sh} + HP_{el} + HP_{hd} \tag{10}$$

$$MB_t = HP_M + UB_t \tag{11}$$

$$LB_t = MB_t + HP_{kn} + HP_{ft} \tag{12}$$
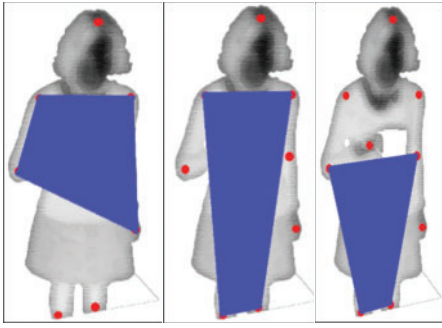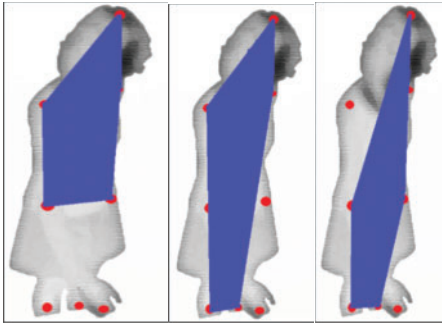
3.4 Angular-Geometric Feature

The angular-geometric feature is a body point feature that extracts shape-based entity features of human skeleton point modeling [32]. The extreme points identified on the human silhouette has been extracted including head, feet, arms, and shoulders. Then, extreme points have been joined together to form three geometric shapes of triangle, quadrilateral, and pentagon. Moreover, changes in the angles have been measured at each extreme point within sequential frames [33]. The geometric shapes have been prepared made by joining the extreme points on the body including inter-silhouette triangle, quadrangular, and pentagon, as shown in Tab. 1.

Table 1: The geometrical shapes of a human silhouette

Type of geometrical shape	Connected extreme points	No. of angles	Diagrammatical representation
Triangle	H + RS + LS H + RA + LA H + RF + LF	9	

(Continued)

Table 1: Continued

Type of geometrical shape	Connected extreme points	No. of angles	Diagrammatical representation
Quadrangular	RS + LS + RA + LA RA + LA + RF + LF RS + LS + RF + LF	12	
Pentagon	H + RS + LS + RA + LA H + RA + LA + RF + LF H + RS + LS + RF + LF	15	
Total 3 types	9 geometrical shapes	36 angles	

Note: H = Head, RS = Right Shoulder, LS = Left Shoulder, RA = Right Arm, LA = Left Arm, RF = Right Foot, LF = Left Foot.

The extreme body points have been joined together to form inter-silhouette geometric shapes. Later on, cosine angle within the shapes have been measured as shown in Eq. (13):

$$\theta = \cos^{-1} \frac{u.v}{|u||v|} \tag{13}$$

where u and v are the vectors that have been used to measure the angle. Next, the area of a triangle is measured using Eq. (14):

$$A_t = \sqrt{S(S - a)(S - b)(S - c)} \tag{14}$$

where a, b, and c are the three sides of a triangle i.e., extreme points, and S depicts semi-perimeter of the triangle.

The silhouette movement may result in the increase or decrease of each geometric shape. Hence, geometric features and angles also variate between sequential frames. The differentiation in angle is higher in a few moments namely, open, pull, pick as they involved rapid movement, than others movements, namely drink, plug, and switch because include less pronounced movement.

3.5 Optimization Using PBIL

Population Based Incremental Learning (PBIL) is a stochastic search optimization technique that converges to the optimal solution based on the learning of current excellent individuals. PBIL algorithm terminates automatically when three of its parameters including learning rate, population size, and search rate converge on a single optimal solution [34]. First, the feature vectors obtained in the Sections 3.2 and 3.4 have been coded with a binary chromosome of population length to form a single chromosome.

The prototype vector P has been generated to bias the generation of chromosomes population [35]. The P with a length equal to the population array has been initialized to 0.5 for each corresponding location to bias the generation of bits. For each chromosome, the bits have been randomly set in the range [0, 1]. The chromosome has been set to either one in case the corresponding random number in the vector P is less than a certain threshold or zero. Next, chromosomes have been evaluated with Eqs. (15)–(16) to incorporate the best chromosome.

$$P_{n+1} = ((1-l)(P_n + l)C_B)(1-f) + \frac{f}{2}(1) \qquad (15)$$

$$f = \frac{2sl}{1 - 2s(1-l)} \qquad (16)$$

where C_B shows best chromosomes, l represents learning rate, and s represents search rate. The algorithm runs iteratively until all chromosomes converge to optimal solution with a convergence level of s or 1-s, see Eq. (17).

$$max(0.5 - |P - 0.5|) < s + \frac{(0.5 - s)}{10} \qquad (17)$$

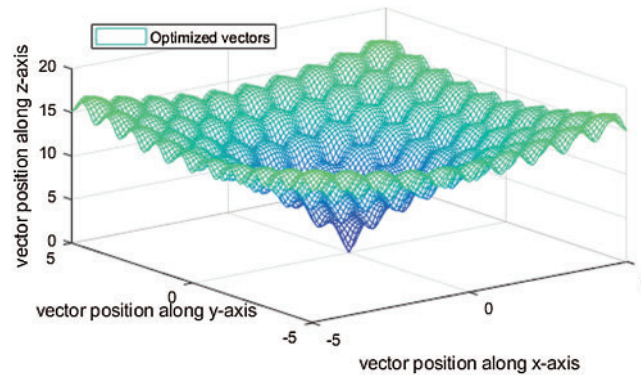Fig. 6 shows the optimization result of probability based incremental learning (PBIL).

Figure 6: 3D graphical representation of PBIL optimization in visual context

3.6 Classification Using Random Forest Classifier

The results of the optimized features have been then fed to the random forest classifier in order to classify features into different activities. Random forest is an ensemble learning algorithm for data classification [36].

Initially, a random forest algorithm has been formed ample number of impartial decision trees to form a forest. The random forest has been then supervised learning impact by assigning the score

to each feature. The decision trees have been then selected for classification that contained abundant number of votes among other decision trees in the whole forest [37]. Eq. (18) illustrates mathematical expression for random forest classifier.

$$Pred = \frac{1}{D} \sum_{d=1}^{D} t_d(x) \tag{18}$$

where x represents predictions of unseen samples. To outfit high dimensional data, the bootstrapping aggregation methodology has been used in the random forest classifier. Hence, the best possible classification tree has been obtained by combining multiple trees together and then fit the training data into regression trees. Finally, all the samples have been employed with replacement to achieve consistency and maximum possible accuracy. Fig. 7 shows visualization clusters of a random forest classifier.
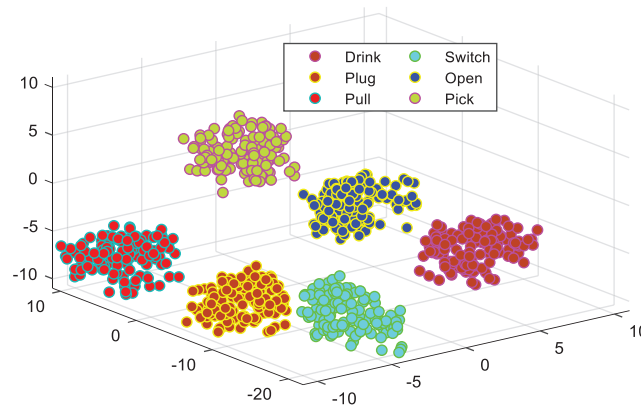

Figure 7: 3D visualization clusters of random forest to segment ADL activities

4 Experimental Setting and Results

In the experimental section, the n-fold cross-validation scheme has been used to evaluate the performance of proposed model over RGBD-AC benchmark datasets. The description of the dataset is also given in this section. Furthermore, the proposed model has been assessed with recognition accuracy, number of precision, observations, recall, computation time, F1 score, and number of states. Finally, the proposed model has been compared against other state-of-the-art methods.

4.1 RGBD-AC Datasets Description

The RGBD-AC dataset [38] embodies RGB and depth images data, taken from optical measuring devices like Microsoft Kinetic v2. This RGBD-AC dataset is divided into 414 sequences of complete and incomplete six different actions. The RGB folders contain RGB sequences of images related to six different actions i.e., RGB-drink, RGB-open, RGB-pick, RGB-plug, RGB-pull, and RGB-switch. Similarly, each depth folders contain depth sequences of images related to six different actions including depth-drink, depth-open, depth-pick, depth-plug, depth-pull, and depth-switch. Each action has been cataloged with 383 images. In this dataset, action has been completed if the goal was achieved i.e., the plug action is completed when a person plugs the switch. In the proposed model, the main focus is on the accuracy for action detection rather than on the action completeness.

4.2 Performance Parameters and Evaluations

The proposed ADL methodology has been validated with the numerous parameters i.e., recognition accuracy, number of precision, observations, recall, computation time, F1-score, and number of states. Details of parameters for individual experiments have been deliberated as:

4.2.1 Experiment I

In the first experiment, the n-fold cross-validation method has been used to evaluate the benchmark dataset for the average accuracy of the proposed model. Tab. 2 shows the confusion matrix for dataset and the mean accuracy of dataset is 92.71%.

Table 2: Confusion matrix showing accuracies of RGBD-AC dataset

Interaction classes	Drink	Open	Pick	Plug	Pull	Switch
Drink	**90.8**	0	0.2	4.5	3.0	1.5
Open	3.04	**92.06**	1.2	2.2	0.5	1.0
Pick	0.1	2.2	**95.9**	0.8	0.5	0.5
Plug	2.5	0	0.5	**93.5**	0	3.5
Pull	0.4	0.3	4.5	1.0	**92.3**	1.5
Switch	2.5	3.5	0	2.0	0.3	**91.7**
Mean Recognition Accuracy = 92.71%						

By means of the proposed effective model, the pick activity has been achieved an accuracy of 95.9%. While some vectors of the plug and switch have been interleaved with each other. Some percentage of drink, open, and pull are also interleaved. It is difficult for the system at some points to differentiate between drink/open and plug/pull due to the repetition of similar movements involved in the activities. However, the overall accuracy of different actions proved the robustness of the proposed model by achieving higher mean accuracy of 92.71%.

4.2.2 Experiment II

The proposed system has been evaluated with an n-fold cross-validation scheme and dividing the dataset into training and testing sets. The results of the PBIL optimizer and random forest classifier have been validated using precision, recall, and F1-score to identify ADL activities. Moreover, the evaluation of RGBD-AC over three parameters have been shown in Tab. 3.

It is observed from Tab. 3 that the drink and switch activities have the least precision rate due to the maximum percentage of false positive. This is due to the body movement of silhouette performed during drink and switch activities involving many actions quite similar to interactions with each other.

Table 3: Comparison of precision, recall and F1 score of the dataset

Dataset	Interactions	Precision	Recall	F1-score
RGBD-AC	Drink	0.90	0.91	0.91
	Open	0.92	0.92	0.93
	Pick	0.94	0.95	0.92
	Plug	0.93	0.94	0.94
	Pull	0.92	0.93	0.94
	Switch	0.91	0.91	0.91
	Average	92.0	92.7	92.5

4.2.3 Experiment III

In the third experiment, various experiments has been performed for the dataset. In this experiment, different sequences of observations has been fused with different states to estimate the accuracy of proposed system based on recognition accuracy and computational time. The performance of the model has effected with the number of states as the current state influenced the performance of previous states. Moreover, the greater the number of observations and states, the greater will be accuracy rate as shown in Tab. 4.

Table 4: Human body key point's detection accuracy

Parameters		Performance	
Number of states	Observations	Computational time (sec)	Accuracy (%)
4	$X = 10$	17	90.0
	$X = 20$	22	90.3
	$X = 30$	23	90.5
5	$X = 10$	15.5	91.6
	$X = 20$	21	91.6
	$X = 30$	26.7	92.1
6	$X = 10$	22.4	92.4
	$X = 20$	25.6	92.5
	$X = 30$	31.7	92.7

From Tab. 4, starting with four states and by varying the observations from 10 to 30 has gradually improved the accuracy rate. This accuracy has been further enhanced in states 4 and 5. Finally, in state 6 the computational time will remained static with no major change in accuracy.

4.2.4 Experiment IV

In fourth experiment, the proposed model has been compared with the other proposed scheme. The Tabs. 5 and 6 shows that the proposed model has achieved a significance performance of 92.7% than other proposed methods.

Table 5: Comparison of proposed method with other significance model over RGBD-AC dataset

Dataset	Authors	Methodology	Recognition accuracy (%)
RGBD-AC	Miron et al. [39]	Convolution neural network	71.2
	Heidarivincheh et al. [40]	HMM and LSTM	75.0
	Heidarivincheh et al. [41]	Discriminative frame-level features	84.0
	Heidarivincheh et al. [42]	Joint-classification regression	89.0
	Proposed methodology	Skeleton point and angular geometric features	**92.7**

Table 6: Comparison of proposed method with other state-of-the-art model

Research area	Authors	Methodology	Recognition accuracy (%)
Human activity recognition	Lee et al. [9]	Skeleton joint features + CNN	70
	Si et al. [12]	Spatial and temporal features	85
	Mahmood et al. [10]	WHITE STAG model	87
	Mahmood et al. [10]	Hour glass network	90.8
	Proposed methodology	Skeleton point and angular geometric features + random forest	**92.7**

5 Discussion

Practical Discussion:

The experiments have been performed on a Desktop PC equipped with Intel(R) Process(R) CPU G3260, 8 GB RAM, 3.30 GHz processing power, 64-bit operating system, Windows 10, and using Anaconda Spyder Python IDE. Our proposed framework has been based on RGB and depth images which help to attain the information of color, appearance, and illumination information

simultaneously. The fused information of RGB and depth has helped to gain the high accuracy and F1-Score. Initially, different preprocessing methodologies has been used to do filtration on RGB-D images to eliminate the noise from raw images of RGB and depth images. Furthermore, full body features and point based features has helped the proposed system in obtaining better performance. Next, attained features have been optimized through Probability based incremental learning. Finally, random forest classifier has been applied to evaluate the performance of the proposed system. The leave-one-subject-out cross-validation (LOSO) method has been used to evaluate the performance of our proposed model on RGB-D benchmark dataset. Moreover, the proposed AAL has been authorized over precision, recall, and F-measure parameters to assess the validity of our model. Finally, a comparison with other state-of-the-art methods has been done to prove that our model out-performed in its best accuracy rate. The results have been showed in Fig. 8.
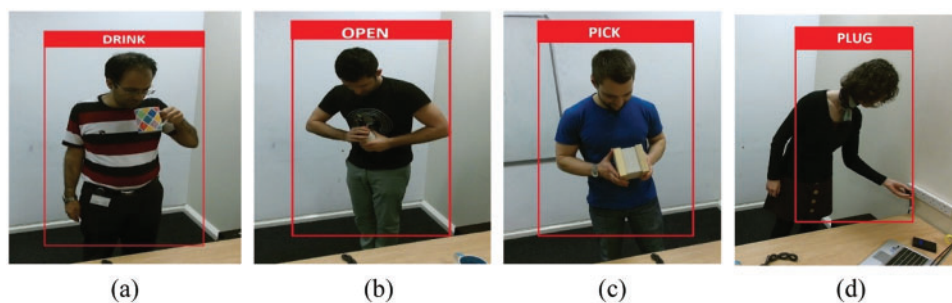

Figure 8: Action recognition of RGBD-AC dataset: (a) Drink, (b) Open, (c) Pick, and (d) Plug

The following are the limitations of our systems.

1. The proposed methodology has been evaluated only on one benchmark dataset which is based on indoor activities. The different datasets could be considered based on diverse patterns of activities and with both indoor and outdoor scenarios for the better performance of our model.
2. The proposed methodology is comprised of full-body and point based features in combination with optimization and classification required high computational time to process the whole results but has achieved the high accuracy of 92.71%.
3. The main problem in PBIL algorithm is the early convergence to the local optimal solution which made it difficult to obtained the excellent results in drink and switch activity.

Theoretical Discussion:

From the obtained results in Tab. 2, the drink activity has obtained an accuracy of 90.8, while rest of the results are confused with pick, plug, pull, and switch activities. Similarly, open, pick, plug, pull, and switch activities have obtained an accuracy of 92.06%, 95.9%, 93.5%, 92.3%, 91.7%, and rest of their results are confused with other activities. The drink activity has obtained best accuracy of 90.8% and pick has obtained the high accuracy of 95.9%. The open and pull activity has obtained almost same accuracy of 92%. While, switch activity has obtained the average accuracy of 91.7%. It can be concluded from the experimental results that combination of features along with optimization and classification delivers better results for the proposed ADL system.

6 Conclusion

In this work, we have proposed a novel RGB-D based ADL recognition system. The main accomplishments of this research effort include identification of human silhouette, integration of full-body feature with a geometric feature, PBIL optimization, and recognition of activities with random forest classifier. The full body features and point based features analyses the optimal pattern, temporal patterns, repetitive pattern, and invariability in motion and thus help to improve the performance of proposed ADL recognition system. This paper has also done comparison against RGB and depth based research work. During the experiment, RGBD-AC benchmark dataset has been used to evaluate the performance of our proposed model. The proposed system achieved an accuracy rate of 92.7% over the other state-of-the-art methods. The proposed system can be applied to various real-life activities including health care, video indexing, and security monitoring.

In the future, we plan to use point cloud features extraction methodology for the ADL recognition model on the RGBD-AC dataset. We will also use static scenes of RGB-D (2.5D) and vision-based 3D visual data of indoor scenes for the enhancement of ADL activities recognition. We will also test our current system on more ADL-based challenging datasets.

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

References

[1] L. M. Powell, "Aging and performance of home tasks," *Human Factors*, vol. 32, no. 5, pp. 527–536, 1990.

[2] D. J. Magermans, E. K. J. Chadwick, H. E. J. Veeger, "Requirements for upper extremity motions during activities of daily living," *Clinical Biomechanics*, vol. 20, no. 6, pp. 591–599, 2005.

[3] A. Prati, C. Shan and K. I. -K. Wang, "Sensors, vision and networks: From video surveillance to activity recognition and health monitoring," *Journal of Ambient Intelligence and Smart Environments*, vol. 11, no. 1, pp. 5–22, 2019.

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

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

[6] F. Farooq, A. Jalal, and L. Zheng, "Facial expression recognition using hybrid features and self-organizing maps," in *Proc. IEEE Int. Conf. on Multimedia and Expo*, Hong Kong, China, pp. 409–414, 2017.

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

[8] M. Mahmood, A. Jalal and H. A. Evans, "Facial expression recognition in image sequences using 1D transform and Gabor wavelet transform," in *IEEE Conf. on Int. Conf. on Applied and Engineering Mathematics*, Taxila, Pakistan, pp. 1–6, 2018.

[9] J. Lee and B. Ahn, "Real-time human action recognition with a low-cost RGB camera and mobile robot platform," *Sensors*, vol. 20, no. 10, pp. 1–12, 2020.

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

[11] N. Varshney, B. Bakariya and A. K. S. Kushwaha *et al.,* "Rule-based multi-view human activity recognition system in real time using skeleton data from RGB-D senso," *Soft Comput.*, vol. 210, pp. 1–17, 2021.

[12] C. Si, Y. Jing, W. Wang, L. Wang, T. Tan, "Skeleton-based action recognition with hierarchical spatial reasoning and temporal stack learning network," *Pattern Recognit.*, vol. 107, pp. 107511–107532, 2020.

[13] S. -T. Kim, H. J. Lee, "Lightweight stacked hourglass network for human pose estimation," *Appl. Sci.*, vol. 10, no. 6497, pp. 1–15, 2020.

[14] D. Madhuranga, R. Madushan and C. Siriwardane, "Real-time multimodal ADL recognition using convolution neural networks," *Vis Comput.*, vol. 37, pp. 1263–1276, 2021.

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

[16] K. Buys, C. Cagniart, A. Baksheev, T. D. Laet, J. D. Schutter *et al.,* "An adaptable system for RGB-D based human body detection and pose estimation," *Journal of Visual Communication and Image Representation*, vol. 25, no. 1, pp. 39–52, 2014.

[17] J. Fu, C. Liu, Y. Hsu and L. Fu, "Recognizing context-aware activities of daily living using RGBD sensor," in *2013 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, Tokyo, Japan, pp. 2222–2227, 2013.

[18] A. Jalal and S. Kamal, "Improved behavior monitoring and classification using cues parameters extraction from camera array images," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 2, pp. 71–78, 2018.

[19] K. Alzahrani and M. Alnfiai, "Evaluation of NFC-guidable system to manage polypharmacy in elderly patients," *Computer Systems Science and Engineering*, vol. 41, no. 2, pp. 445–460, 2022.

[20] S. Abbas, Y. Alhwaiti, A. Fatima, M. A. Khan, T. M. Ghazal *et al.,* "Convolutional neural network based intelligent handwritten document recognition," *Computers, Materials & Continua*, vol. 70, no. 3, pp. 4563–4581, 2022.

[21] Y. Guo, Z. Cui, X. Li, J. Peng, J. Hu *et al.,* "MRI image segmentation of nasopharyngeal carcinoma using multi-scale cascaded fully convolutional network," *Intelligent Automation & Soft Computing*, vol. 31, no. 3, pp. 1771–1782, 2022.

[22] A. Jalal, M. A. K. Quaid and M. A. Sidduqi, "A triaxial acceleration-based human motion detection for ambient smart home system," in *IEEE Int. Conf. on Applied Sciences and Technology*, Islamabad, Pakistan, pp. 353–358, 2019.

[23] A. Jalal, M. Mahmood and A. S. Hasan, "Multi-features descriptors for human activity tracking and recognition in indoor-outdoor environments," in *IEEE Int. Conf. on Applied Sciences and Technology*, Islamabad, Pakistan, pp. 371–376, 2019.

[24] Q. Ye, H. Zhong, C. Qu and Y. Zhang, "Human interaction recognition based on whole-individual detection," *Sensors*, vol. 20, no. 8, pp. 1–18, 2020.

[25] O. Ouyed and M. A. Said, "Group-of-features relevance in multinomial kernel logistic regression and application to human interaction recognition," *Expert Systems with Applications*, vol. 148, pp. 1–22, 2020.

[26] A. A. Rafique, A. Jalal and A. Ahmed, "Scene understanding and recognition: Statistical segmented model using geometrical features and Gaussian naïve Bayes," in *IEEE Conf. on Int. Conf. on Applied and Engineering Mathematics*, pp. 225–230, 2019.

[27] M. Batool, A. Jalal and K. Kim, "Sensors technologies for human activity analysis based on SVM optimized by PSO algorithm," *IEEE ICAEM Conf.*, Taxila, Pakistan, pp. 145–150, 2019.

[28] A. Shehzad, A. Jalal and K. Kim, "Multi-person tracking in smart surveillance system for crowd counting and normal/abnormal events detection," in *IEEE Conf. on Int. Conf. on Applied and Engineering Mathematics*, Taxila, Pakistan, pp. 163–168, 2019.

[29] A. Ahmed, A. Jalal and A. A. Rafique, "Salient segmentation based object detection and recognition using hybrid genetic transform," *IEEE ICAEM Conf.*, Taxila, Pakistan, pp. 203–208, 2019.

[30] S. Bibi, N. Anjum and M. Sher, "Automated multi-feature human interaction recognition in complex environment," *Computers in Industry Elsevier*, vol. 99, pp. 282–293, 2018.

[31] Y. Ji, H. Cheng, Y. Zheng and H. Li, "Learning contrastive feature distribution model for interaction recognition," *Journal of Visual Communication and Image Representation*, vol. 33, pp. 340–349, 2015.

[32] J. Peng, C. Xia, Y. Xu, X. Li, X. Wu *et al.*, "A multi-task network for cardiac magnetic resonance image segmentation and classification," *Intelligent Automation & Soft Computing*, vol. 30, no. 1, pp. 259–272, 2021.

[33] M. u. Rehman, S. H. Khan, S. M. Danish Rizvi, Z. Abbas, A. Zafar *et al.,* "Classification of skin lesion by interference of segmentation and convolotion neural network," in *2nd Int. Conf. on Engineering Innovation (ICEI)*, Bangkok, Thailand, pp. 81–85, 2018.

[34] A. Ahmed, A. Jalal and K. Kim, "RGB-D images for object segmentation, localization and recognition in indoor scenes using feature descriptor and Hough voting," in *IEEE Conf. on Applied Sciences and Technology*, Islamabad, Pakistan, pp. 290–295, 2020.

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

[36] M. A. K. Quaid and A. Jalal, "Wearable sensors based human behavioral pattern recognition using statistical features and reweighted genetic algorithm," *Multimedia Tools and Applications*, vol. 79, pp. 6061–6083, 2019.

[37] A. Nadeem, A. Jalal and K. Kim, "Human actions tracking and recognition based on body parts detection via artificial neural network," in *IEEE Int. Conf. on Advancements in Computational Sciences*, Lahore, Pakistan, pp. 1–6, 2020.

[38] F. Heidarivincheh, M. Mirmehdi and D. Damen, "Beyond action recognition: Action completion in RGB-D data," in *27th British Machine Vision Conf.*, York, United Kingdom, 2016.

[39] A. Miron and C. Grosan, "Classifying action correctness in physical rehabilitation exercises," in *IEEE Conf. on Int. Conf. on Applied and Engineering Mathematics*, Stockholm, Sweden, 2021.

[40] F. Heidarivincheh, M. Mirmehdi and D. Damen, "Detecting the moment of completion: Temporal models for localising action completion," in *28th British Machine Vision Conf.*, London, UK, 2017.

[41] F. Heidarivincheh, M. Mirmehdi and D. Damen, "Weakly-supervised completion moment detection using temporal attention," in *30th British Machine Vision Conf.*, Seoul, South Korea, 2019.

[42] F. Heidarivincheh, M. Mirmehdi and D. Damen, "Action completion: A temporal model for moment detection," in *29th British Machine Vision Conf.- Northumbria University*, Newcastle upon Tyne, United Kingdom, 2018.