IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received June 9, 2021, accepted June 20, 2021, date of publication June 25, 2021, date of current version July 5, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3092403

A Unified Approach for Patient Activity Recognition in Healthcare Using Depth Camera

MUHAMMAD HAMEED SIDDIQI[1], NABIL ALMASHFI[1], AMJAD ALI[2], MADALLAH ALRUWAILI[1], YOUSEF ALHWAITI[1], SAAD ALANAZI[1], AND M. M. KAMRUZZAMAN[1], (Senior Member, IEEE)

[1]College of Computer and Information Sciences, Jouf University, Aljouf 72311, Saudi Arabia
[2]Department of Computer Science, COMSATS University Islamabad, Lahore Campus, Lahore 54000, Pakistan

Corresponding author: Muhammad Hameed Siddiqi (mhsiddiqi@ju.edu.sa)

This work was supported by the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia, under Project 375213500.

This work did not involve human subjects or animals in its research.

ABSTRACT Context-awareness is an essential part of pervasive computing. Video-based human activity recognition (HAR) has arisen as an imperative module to detect user's situation for involuntary facility delivery in context-aware domains. The activity recognition systems are frequently employed for protective and practical health care. Most of the existing works utilize RGB (red, green, and blue) cameras which present confidentiality and security concerns in the health-care domain. The existing approaches also do not sustain their performance results under the presence of a depth camera. Moreover, the accuracy of an HAR system relies on the extraction and selection of the prominent features from the feature space. To address these limitations, in this research, we first employ a depth camera to resolve the confidentiality and security concerns, and propose an unsupervised segmentation algorithm that can accurately segment the human body from the video frame. Then, we propose a new feature selection technique that has the ability to excerpt and select the best features from the feature space. Our proposed feature selection method focuses on the selection of confined features from the series of images and discriminate their category based on reversion (i.e., regression) value. The proposed method extracts and selects the best features through the benefits of forward selection and backward regression algorithms. Finally, we have trained and tested our proposed system by employing hidden Markov model (HMM) to label the activities. The proposed approach presents a significant performance against the existing works using depth camera.

INDEX TERMS Healthcare, activity recognition, depth camera, segmentation, level set, contour, SWLDA, HMM.

I. INTRODUCTION

CONTEXT awareness is at the heart of pervasive computing. It is the user context that helps an application everywhere automatically to deliver the right kind of service such as making an emergency call to the clinic in the event of an elderly patient falling or turn off the lights when a user leaves their office/home.

Activity recognition is a key element in defining a user's context for delivering services based on the kind of ubiquitous tools. For instance, in the case of ubiquitous care applications, knowledge of day-to-day activities may enable these systems to see and learn changes in an elder's daily behavior that may be indicative of their health progress e.g. patient's behavior [1], monitoring of the patients [2], fight detection [3]. In addition, it can help determine the degree of independence of the elderly, understand the side effects of medication and promote compliance.

Video activity recognition is often explained as automatic recognition of human physical activities by computer using depth cameras. The accuracy of these systems depend largely on the extraction and selection of segments of human body, and hence on the performance of segmentation process. The reason why the segmentation of the human body is so

The associate editor coordinating the review of this manuscript and approving it for publication was Byung-Gyu Kim.

important in such systems is that it defines the necessary and sufficient area in an image for other modules to focus such as feature extraction. In other words, it defines the body shapes used to extract the features, which directly helps adds to the accuracy of the recognition process. Many works have been presented which tried to define the contours of the human body (body and object are utilized interchangeably in some parts) simply by performing a low-level image processing task, such as canny edge detection. This, however, is not a technical solution. It is particularly effective when the edges are sharp and generally not incessant. It exists because of the noise [4].

In general, a HAR system has three common elements: segmentation, feature extraction & selection, and recognition. In essence, human bodies can perform diverse set of activities, and hence possess a lot of activity information. That is why in the segmentation element, the human body is first segmented and divided in a specific context. The feature extraction and selection part deals with the acquisition and valuation of the distinguishing features of any kind of activity as a token. In the recognition part, first a classifier is skilled on the training data and then it is used to generate appropriate activity tags contained in the received video data.

In most of the existing works, binary silhouettes are extensively applied to symbolize a variety of body configuration. Binary silhouettes sometimes generate uncertainties by presenting the similar shape for various postures from various actions. For example, if someone makes a natural movement of hand in front of the camera, different postures might correspond to the same shapes due to the binary plane and the flat distribution of pixel density [5]. In such cases, binary shapes do not seem selected the preferred choice for distinguishing these various positions from different activities. Figure 1 shows the RGB binary and depth activities such as hands up and down, clapping, and boxing individually. Obviously, binary shapes are a wrong choice for each separate various posture. Unlike binary silhouettes, where there is a flat binary distribution of the pixels [5], in depth silhouettes body pixels are dispersed based on the distance to the camera, making them a superior choice over binary silhouettes. The depth silhouettes can be obtained by using the infrared sensor-based depth camera or disparity calculation of the pixels in the stereo RGB images captured using a stereo camera [6].

Just like RGB-camera based techniques, such as [5]–[7], some of the existing works that use depth-camera such as [8]–[10] segment the human body from depth just by subtracting the empty frame from depth video frame. Due to this, these methods are also known as heuristic techniques.

Ubiquitous health-care everywhere rises privacy issues as it may lead to circumstances where patients are unaware that their confidential data is being communal and therefore could be at risk [11]. Most of the ubiquitous health-care systems developed in the past employed RGB camera to monitor patients' daily routine and detect their activities. These are some common concerns about privacy that RGB cameras have in such systems. Therefore, in this case, depth-cameras are a good candidate to maintain patients' privacy.

Existing systems [12]–[17] fail to achieve best accuracy due to week and heuristic feature extraction and selection techniques. The results of majority of existing works are presented in controlled environments. The performances of feature extraction and selection methods employed in their (existing works) respective systems degrade because of dynamic activities and controlled environments. Therefore, in this research, we develop an intelligent HAR system against depth camera that considers most of the limitations of the existing work. Firstly, we present a vigorous human body segmentation algorithm, which is an unsupervised method that efficiently perceives and segments the human body from the video frame. The proposed segmentation algorithm is the association of two energy functions (Chan-Vese and Battacharya functions) which not only reduces the similarities within the human body parts but also increases the detachment between the human body and the background [18]. Our proposed segmentation technique shows utmost performance on static and dynamic activities (which is one of the limitations of our previous work [19].

Once a human body in a frame/picture is segmented, then we apply our proposed novel feature extraction and selection approach that extracts and selects the utmost features from the activity frames. The developed method concentrates on the collection of confined features from the series of frames and discriminates their class based on regression values. The proposed technique has the ability to simultaneously select the best features and remove the redundant features via forward and backward models, respectively. Finally, we make use of hidden Markov model to name the activities and assess the performance of the proposed system by employing a set of eleven activities that include running, walking, jumping, skipping, one hand, two hand waving, bending, place jumping, side movement, clapping, and boxing. These activities are performed by 50 subjects (university students) and captured using a depth camera.

The rest of the paper is organized as: Section II summarizes state-of-the-art activity recognition systems using depth camera. Section III presents the proposed model. The dataset which is utilized in this work is described in Section IV. The experimental setup for the proposed model is presented in Section V. The results with discussion are presented in Section VI. Finally, the paper will be concluded with some future directions in Section VII.

II. RELATED WORKS

Commonly, an HAR system consists of three basic units i.e. segmentation, feature extraction/selection, and recognition. Plenty of research material exist for activity recognition, however, a limited amount of work exists in the fields of segmentation and feature extraction/selection. This is mainly because of the difficulty in selecting prominent features from the feature space.

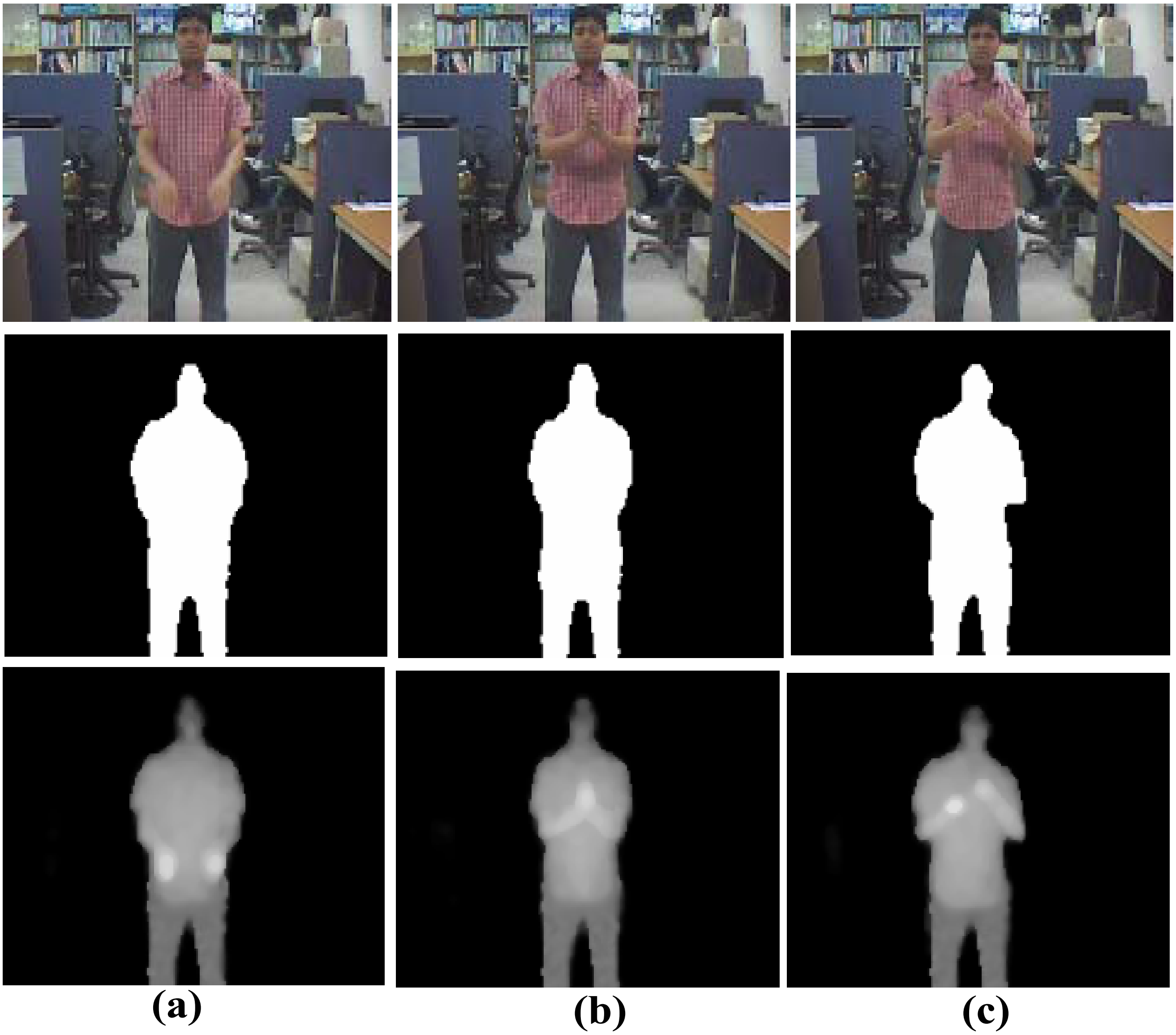

FIGURE 1. **Different types of human activities. (a) shows both hands up and down, (b) represents hand clapping and (c) indicates boxing activities respectively [5]. It is obvious from the figure that the binary silhouettes cannot provide good features through which we can discriminate such types of activities, but on the other hand depth frames provides best features for recognizing these activities easily.**

An integrated approach proposed by [20] for the purpose of segmenting the human body from the video frame utilizes a model base technique that detects the human body through the combination of 2D head contour method and 3D head surface model. However, these methods do not have the ability to present the real time objects through logical surfaces [21]. Similarly, a real time system proposed by [22] investigates the possibility to segment the foreground (human body). They utilize SDK along with sensors permits to get the effective detection of the foreground human bodies but exclude generic objects. However, SDK has a limitation that it cannot support large images i.e. having bigger sizes. It means that their accuracy degrades when employed against sequence of images. Moreover, a real time body segmentation method is proposed in [23]. In this real-time body segmentation method, superparamagnetic clustering of data for segmentation is used which determines the equipoise of the states of a Potts model. Superparamagnetic clustering algorithm shows better performance in frame-based classification; however, its accuracy degrades when applied on sequential data [24]. The authors of [25], [26] developed state of the art methods respectively for trajectory classification (using segmental HMM) and facial expression recognition (using deep learning), which might be employed for activity recognition.

Similarly, large numbers of publications exist on feature extraction and selection; however, most of them have their own limitations. For example, a hybrid technique proposed by [13] for the purpose of feature extraction integrates optical flow-based motion features with distance factors features in order to extract the prominent features from the depth body. Then, utilize those in an augmented form in order to work as spacio-temporal features. Generally speaking, a feature-based method has no ability to yield a more precise and thorough 360-panorama. This is because the depth of the projected region and the location of the center point are dissimilar between cameras conferring to the location and direction in head-to-head cameras [27]. [28], [29] are on feature selection; however, [30] is on random projection (RP) for dimensionality reduction, similar to [31], [32]. [28], [31], [32] are on Two-Directional Two-Dimensional RP (2D–2RP) and Two-Dimensional RP (2DRP), they are less expensive computationally and required lower storage than traditional one dimensional random projection (1DRP) [33]. A naturalistic fusion system is developed by [34] in order to recognize the human actions, which employs data from depth camera and inertial sensor. The developed system combines the probability productions of the features from such modalities in real environment through a decision-based fusion technique that is further fed to collaborative classifier. However, it is very much difficult to sustain the system an admirable classification result [35], [36].

In a more recent work, the authors of [37] proposed a hybrid feature selection method that is an integration of a filter and wrapper methods. They claim a higher recognition rate; however, they have utilized smart-phone for activity recognition, which has limitations in real world environments due to signal loss and storage that may cause misclassification. Similarly, the authors of [38] proposed a novel integrated algorithm for feature extraction and selection based on Magnitude of the Signal, Naive Bayes, k-Nearest Neighbor and Random Forest. They achieved better accuracy than others using iPhone and other accelerometer-based devices. However, k-Nearest Neighbor is a court-based learning algorithm, which is also known as indolent learner because it records all training samples and do not create a classification until a new anonymous sample is examined [39], [40], and random forest is computationally way expensive as compared to others. In another work, the authors propose ensemble-based filter feature selection (EFFS) for human activity recognition by [41]. The authors in this work show better performance as compared to other approaches; however, most of ensembled-based feature selection techniques do not take into account the benefits of certain characteristics, as explained in [42].

Examples of these characteristics include reducing feature sample space size by exporting special features based on the idea of maximizing data spread and reducing category variation. The characteristic value for activity classes has been confirmed to be very small, which can lead to a high level of misclassification. This is because of the similarity among the activities that cause high class variation and low-class variation. Therefore, the required method does not only reduce the size, but also increases the small variation between classes to increase the separation of classes before adding features to the classifier.

Furthermore, some machine learning algorithms [43]–[45] have been developed in order to resolve the above-mentioned problem. Of these, linear discrimination analysis (LDA) has been extensively used in activity recognition systems. However, LDA has two main concerns. First, it is based on a mixed model that contains the correct amount of ingredients. Second, this is a linear algorithm that has limited suppleness when using the same table. Also, LDA assumes that all classes have the same environmental table results in a strong LDA transformation and there may not be enough data to estimate the lags in class separation. For more information on LDA, see the previous study [46].

In this work, we have proposed an intelligent HAR system in which we propose an unsupervised human body segmentation algorithm and a new non-linear feature extraction and selection method. The segmentation algorithm is based on the association of two energy functions which not only reduce the variations withing the object (like objects in a human body) but also enlarges the distance amongst the body and background. While, the proposed feature extraction and selection algorithm only focus on local features where it extracts and selects the utmost features from the feature space by taking the advantages of forward selection and backward regression algorithms. The system is further trained and tested utilizing hidden Markov model to name the corresponding activity.

III. MATERIALS AND METHODS

The overall workflow for the video-based HAR is presented in Figure 3.

The detailed components explanation of the developed approaches are described in the following subsections.

A. PROPOSED HUMAN BODY SEGMENTATION ALGORITHM

The first method such as active contour (AC) that was developed by [47] for the purpose of segmentation got much attention for this purpose.

The AC model is the distorted part affected by pressure and force of the image flowing around the boundary of the object. Try to get into a position where the energy is low. Active contours attempt to expand themselves by imposing wanted possessions like steadiness and flatness on the contours of human bodies that means that the active contour focus adds to the level of prior knowledge in solving contour problems in order to find the outline of the object.

Lately, Chan-Vese (CV) proposed in [48] a new procedure of AC for object segmentation rely on level set context. For which the energy function is described as.

$$F(C) = \int_{inside(C)} |I(x) - c_{in}|^2 dx + \int_{outside(C)} |I(x) - c_{out}|^2 dx \tag{1}$$

where c_{in} and c_{out} are the average intensities in the curve of the variable C, respectively. Associated to other AC models,

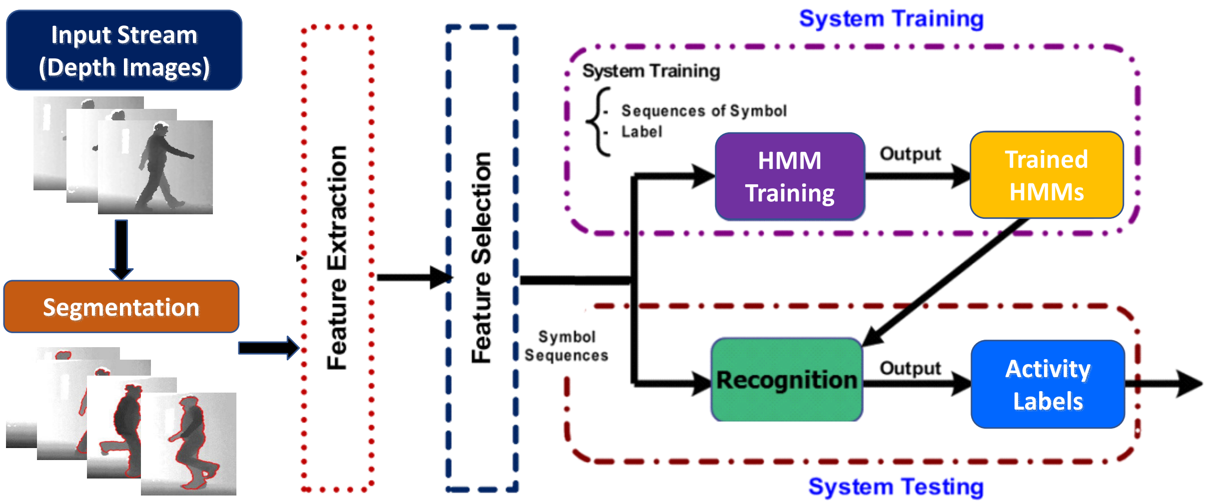

FIGURE 2. The workflow for the generalized video-based HAR system.

the CV approach can see the human body more accurately, as it is not necessary to smooth the image of the original body (using $g\,|\nabla I_\sigma|^2$), even in other words, this model is more resistant to noise. However, the concentration of the CV AC approach relies on the homogeneity of the segmented body, and when the in-homogeneity becomes high, the CV AC approach gives unsatisfactory results. Unlike other contour models that depend on the slope of the frame as the stopping term and therefore gives an unsatisfactory result in noisy and blur image. The fault of the CV AC approach is that the segmentation is only the average value, which means that it is not information but it is useful to differentiate inside and outside curved areas, and makes this is one of the techniques. It is the most powerful and used in s wide way for image segmentation, especially in human body segmentation. The power functions mentioned above do not always guarantee the desired results. The unsatisfactory result of the CV AC approach in this case is due to the fact that it tries to lessen the deviations in each region, but it does not take into account the distance between the different regions.

The proposed methodology is to include terms based on the evolution of Bhattacharyya distance in the CV of energy functions that reduce object differences and increase the distance between two fields. The proposed energy functions are:

$$E_0(C) = \beta F(C) + (1-\beta) B(C) \tag{2}$$

where $\beta \in [0, 1]$. Note that to compare the term $F(C)$, in practice $B(C)$ is multiplied by the area of the body frame, because always its value is in the range [0, 1]; while, $F(C)$ is estimated based on the integration of the activity frame. As traditional [48], we check the solution by limiting the measurement of the curve and the area of the region, thus obtaining the amount of operational energy

$$E(C) = \gamma Length(C) + \eta Area(inside(C)) + \beta F(C) + (1-\beta) B(C) \tag{3}$$

where γ and η are the coefficients of non-negative.

The perception behind the developed approach, in $E(C)$, is that we propose the normal curve (first two terms) and divide the activity frame into zones to minimize the difference in each zone ($F(C)$ term) as a shrink mean Distance between two areas (such as object (human body) and background) increases ($B(C)$ term).

To format the level set ϕ as a function of level set, $I : \Omega \to Z \subset R^n$ as specific image attributes such as density, color, texture, or amalgamation, and $H(\bullet)$ and $\delta_0(\bullet)$ serve as Heaviside and Dirac, correspondingly.

$$H(u) = \begin{cases} 1, & if\ u \geq 0 \\ 0, & if\ u < 0 \end{cases} \qquad \delta_0(u) = \frac{d}{du} H(u) \tag{4}$$

The energy function might be revised as

$$\begin{aligned} E(\phi) = {} & \gamma \int_\Omega |\nabla H(\phi(x))|\, dx + \eta \int_\Omega H(-\phi(x)) \\ & + \beta \begin{bmatrix} \int_\Omega |I(x) - c_{in}|^2 H(-\phi(x)) \\ + \int_\Omega |I(x) - c_{out}|^2 H(\phi(x)) \end{bmatrix} \\ & + (1-\beta) \int_Z \sqrt{p_{in}(z)\, p_{out}(z)} dz \end{aligned} \tag{5}$$

where

$$\begin{aligned} p_{in}(z) &= \frac{\int_\Omega \delta_0(z - I(x)) H(-\phi(x))\, dx}{\int_\Omega H(-\phi(x))\, dx} \\ p_{out}(z) &= \frac{\int_\Omega \delta_0(z - I(x)) H(\phi(x))\, dx}{\int_\Omega H(\phi(x))\, dx} \end{aligned} \tag{6}$$

In universal form, it delivers

$$E(\phi) = \underbrace{\int_\Omega f(\phi, \phi_{x_1}, \phi_{x_2}, \ldots, \phi_{x_n})\, dx}_{\overline{F}(\phi)} + (1-\beta) B(\phi) \tag{7}$$

where $X = [x_1, x_2, \ldots, x_n] \in R^n$, $\phi_{x_i} = \frac{\partial \phi}{\partial_{x_i}}$, $i = \overline{1..n}$, $B(\phi) = \int_z \sqrt{p_{in}(z)\, p_{out}(z)} dz$.

The first difference (w.r.t $\phi(x)$) is described as

$$\frac{\delta E}{\delta \phi} = \frac{\delta \bar{F}}{\delta \phi} + (1-\beta)\frac{\delta B}{\delta \phi} \tag{8}$$

Using Euler-Lagrange equation, one has

$$\begin{aligned}\frac{\delta \bar{F}}{\delta \phi} &= \frac{\partial f}{\partial \phi} - \sum_{i=1}^{n} \frac{\partial}{\partial x_i}\frac{\partial f}{\partial \phi x_i} \\ &= \delta_0(\phi)\left[-\eta - \beta(I-c_{in})^2 + \beta(I-c_{out})^2 - \gamma k\right]\end{aligned} \tag{9}$$

Alternatively,

$$\frac{\delta B}{\delta \phi} = \frac{1}{2}\int_z \left(\begin{array}{c} \frac{\partial p_{in}(z)}{\partial \phi}\sqrt{\frac{p_{out}(z)}{p_{in}(z)}} \\ +\frac{\partial p_{out}(z)}{\partial \phi}\sqrt{\frac{p_{in}(z)}{p_{out}(z)}} \end{array} \right) dz \tag{10}$$

where $p_in(z)$ and $p_out(z)$ are given in (6). Distinguishing them w.r.t $\phi(x)$, one attains

$$\begin{aligned}\frac{\partial p_{in}(z)}{\partial \phi} &= \frac{\partial_0(\phi)}{A_{in}}[p_{in}(z) - \delta_0(z-I)] \\ \frac{\partial p_{out}(z)}{\partial \phi} &= \frac{\partial_0(\phi)}{A_{out}}[\delta_0(z-I) - p_{out}(z)]\end{aligned} \tag{11}$$

where A_{in} and A_{out} are correspondingly the regions inside and outside the curve and are specified by

$$A_{in} = \int_\Omega H(-\phi(x))\, dx \quad A_{out} = \int_\Omega H(\phi(x))\, dx \tag{12}$$

Replacing (11) in (10) and taking some simple alteration, one acquires

$$\frac{\delta B}{\delta \phi} = \delta_0(\phi)\, V(x) \tag{13}$$

where

$$\begin{aligned}V(x) = &\frac{B}{2}\left(\frac{1}{A_{in}} - \frac{1}{A_{out}}\right) \\ &\delta_0(z - I(x)) \\ &+\frac{1}{2}\int_z \left(\frac{1}{A_{out}}\sqrt{\frac{p_{in}(z)}{p_{out}(z)}} - \frac{1}{A_{in}}\sqrt{\frac{p_{out}(z)}{p_{in}(z)}}\right) dz\end{aligned} \tag{14}$$

Merging (8), (9), and (13), one might originate the first distinction of $E(\phi)$ as

$$\frac{\partial E}{\partial \phi} = \delta_0(\phi)\left[\begin{array}{c} -\gamma k - \eta - \beta(I-c_{in})^2 \\ +\beta(I-c_{out})^2 + (1-\beta)V \end{array}\right] \tag{15}$$

Hence, the valuation flow connected with lessening the energy function in (5) is given as

$$\begin{aligned}\frac{\partial \phi}{\partial t} &= -\frac{\partial E}{\partial \phi} \\ &= \delta_0(\phi)\left\{\begin{array}{l} \gamma k + \eta + \beta\left[(I-c_{in})^2 + (I-c_{out})^2\right] \\ -(1-\beta)\left[\begin{array}{l} \frac{B}{2}\left(\frac{1}{A_{in}} - \frac{1}{A_{out}}\right) \\ \quad \delta_0(z-1) \\ +\frac{1}{2}\int_z \left(\begin{array}{c} \frac{1}{A_{out}}\sqrt{\frac{p_{in}}{p_{out}}} \\ -\frac{1}{A_{in}}\sqrt{\frac{p_{out}}{p_{in}}} \end{array}\right) dz \end{array}\right] \end{array}\right\}\end{aligned} \tag{16}$$

Hence, the proposed AC approach tackle the concerns of conventional CV AC approach in the field of segmentation.

B. FEATURE EXTRACTION AND SELECTION

In this section, we discuss the Fisher linear discriminant (FLD) which is very renowned linear classification and used for separation between two classes [43]. Gaussian distribution method can be used for two classes having same covariance, but FLD is more robust classifier in order to find the best separation between the classes. FLD is comparable method with regression methods such as least-square regression method, and also project their feature masses in binary jobs. Proficiency of FLD to show good classification results but only for linear data. We proposed an idea to deal with non-linear classification techniques which has been verified using P300 Speller response [49]. The proposed technique will work in parallel as compared to FLD: reducing feature space and removing inappropriate features.

Moreover, the proposed technique will use the select the best features using two algorithms which work in parallel, namely forward and backward algorithms. The most significant value that has been obtained with the model is "p-value $<$ 0.27" with no primary model at the beginning. After the values entered using forward algorithm, the inappropriate values removed using back projection algorithm such as "p-values $>$ 0.15". This process continues until the predefined condition satisfied and resultant function is constrained to 300 features.

Regression method choose best variables such as X and then move on to make more $X's$ in significant way. The process of adding new entries and the selection of values depend of $F-$test value that determined which entry should be added first or second. After that, comparison is done between two values namely: partial $F-$value and selected value. Forward method is used in this whole process. The deletion process is done using backward regression method (known as backward deletion). In this procedure, the testing that is described in the accumulation are estimated. If the assessment value has least value, then V_L, is associated with the primary selected value. Based on stepwise regression the model is built to show iterations. In each iteration, there is automatic selections of independent variables. The proposed approach will be based on stepwise regression uses both forward and backward regression that include all independent variables and eliminates which are not statistically important from stepwise model [49].

1) PROCESS OF THE PROPOSED APPROACH

Initially, there will be no variables in the developed model. Based on the importance test: For example: partial $F-$test, the projecting variables are passed in or took away from the model at every iteration. Two variables, alpha-to-enter and alpha-to-eliminate, were identified to control the level of importance. Alpha-to-enter = 0.55 and alpha-to-remove = 0.4 as threshold parameters. This level of significance also indicates the importance of the predictor variables introduced or removed from the model. The algorithm stops the itch when the predictions from the progressive model can no longer be inserted or subtracted.

For example, there are three independent variables x_1, x_2, and x_3, and 'y' is denoted as output variable. Regression is used to fit variables into the model. 'p' denotes the number of predictor variables, then regress y on x_1, regress y on x_2 and regress y on x_1. The predictor with the smallest $t-$test along with $P-$value; i.e., below $a_{en} = 0.35$; is entered in the stepwise model first. This process continues until stopping criteria is satisfies; i.e., there are no variables that has $P-$value of least than a_{en}. Consider x_1 is finest predictor. Next, we fit remaining predictor with the finest predictor x_1 in the model; i.e., regress y on (x_1, x_2), regress y on $(x_1, x_3), \ldots, y$ on (x_1, p_{p-1}). In the second step the predictor with the smallest $P-$value ($a_{en} = 0.55$) is inserted into the stepwise model. Again, the iteration stops when there is no $P-$value that is less than 0.55. Consider, x_2 is the "best second predictor" in the model. The algorithm phases back and checks for the value of $P-$value for $\beta_1 = 0$; i.e., the predictor variable removal criteria from the model. The variable is considered as not significant compared to the new entry if the $P-$value has (above $\alpha_{re} = 0.4$) for $\beta_1 = 0$. Contrary to this, let's suppose if both the variables x_1 and x_2 are selected in the two-predictor stepwise approach. Then the algorithm accepts every of the three-prognosticator models through x_1 and x_2 in the model; like regress y on (x_1, x_2, x_3) and regress y on $(x_1, x_2, x_4), \ldots,$ and regress y on (x_1, x_2, x_{p-1}). Third prognosticator that passes in the model is that prognosticator which has the minimum value such as $< \alpha_{en} = 0.55$). When the stopping criterion is met there is no $P-$value $< \alpha_{en}$. In this case when stopping criteria is met algorithm checks back the $P-$value for $\beta_1 = 0$. If the value is greater than $\alpha_{re} = 0.4$), then the prognosticator is eliminated from the model. The procedure breaks after accumulating additional prognosticators that do not yield a $P-$value that is less tha $\alpha_{en} = 0.55$.

C. ACTIVITY RECOGNITION USING HIDDEN MARKOV MODEL (HMM)

The Hidden Markov Model (HMM) is a probabilistic classifier originally proposed by [50] and is often used to identify and classify statistical models. It is an influential tool for noticing a weak signal and is utilized effectively to recognize sequential patterns like speech, handwriting, facial expression and activity recognition. HMM is a statistical approach that employs the Markov procedure along with unseen/hidden and unidentified parameters. The parameters observed in this model are used to represent hidden parameters. This parameter is used for a more detailed analysis. The basic principle of HMM is that the observed events are not the same as individual cases, but refer to cases with probability distributions. This is a multi-random process that includes the Markov chain as the main random process and describes the transition of probabilistic states and processes that illustrate the statistical correspondence between the observed conditions and the values. Unlike the situation, from the point of view of society, only values can be seen. The stochastic process is used for the existence and nature of the case. For this reason, this is named as hidden Markov model. Therefore, for activity recognition, we utilized HMM in order to label the corresponding activities. For more details on HMM, please refer [51].

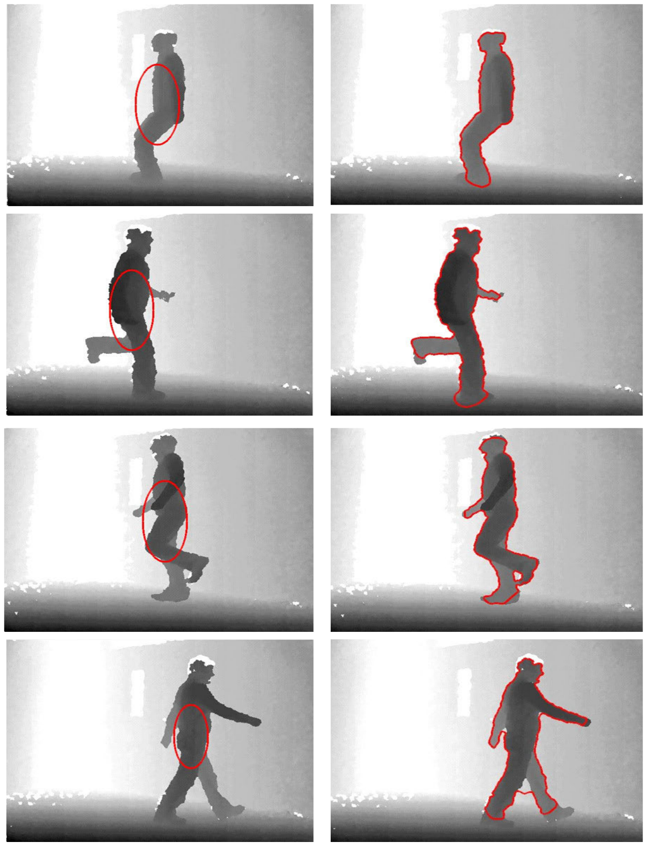

FIGURE 3. Segmentation Results of the proposed model on dynamic activities (like first row is jumping activity (with $\beta = 0.3$, $\gamma = 0.7$, $\eta = 0.0$, and cpu time = 105s), second row is running activity (with $\beta = 0.2$, $\gamma = 0.7$, $\eta = 0.0$, and cpu time = 121s), third row is skipping activity(with $\beta = 0.2$, $\gamma = 0.7$, $\eta = 0.0$, and cpu time = 81s), and the fourth row is walking activity (with $\beta = 0.2$, $\gamma = 0.7$, $\eta = 0.0$, and cpu time = 75s) respectively). Similarly, the first column is the initial contour; while, the second column is final contour.

TABLE 1. The assessment results of the proposed approach (Unit: %). Where WA is used for walking, RU is used for running, JU is used for jumping, SK is used for skipping, OW is used for one hand waving, TW is used for two hand waving, BE is used for bending, PJU is used for place jumping, SMO is used for side movement, CL is used for clapping, and BO is used for boxing.

Classes	WA	RU	JU	SK	OW	TW	BE	PJU	SMO	CL	BO
WA	**96**	0	0	1	0	1	1	0	0	0	1
RU	0	**97**	0	2	0	0	0	1	0	0	0
JU	0	0	**95**	2	0	0	1	0	2	0	0
SK	0	0	0	**98**	0	0	0	2	0	0	0
OW	0	0	0	0	**99**	0	0	1	0	0	0
TW	0	0	0	2	0	**98**	0	0	0	0	0
BE	0	0	1	0	2	0	**96**	0	0	1	0
PJU	0	0	0	0	2	0	0	**98**	0	0	0
SMO	2	0	0	0	0	0	0	0	**97**	0	1
CL	0	0	0	0	0	0	0	3	0	**97**	0
BO	0	0	0	0	0	0	0	0	0	0	**100**
Average						**97.4**					

TABLE 2. The assessment results of the HAR system in the presence of local binary pattern (while, in the absence of the proposed feature extraction and selection method) (Unit: %). Where WA is used for walking, RU is used for running, JU is used for jumping, SK is used for skipping, OW is used for one hand waving, TW is used for two hand waving, BE is used for bending, PJU is used for place jumping, SMO is used for side movement, CL is used for clapping, and BO is used for boxing.

Classes	WA	RU	JU	SK	OW	TW	BE	PJU	SMO	CL	BO
WA	**85**	0	0	0	8	0	0	0	7	0	0
RU	7	**88**	0	5	0	0	0	0	0	0	0
JU	0	0	**90**	5	0	0	5	0	0	0	0
SK	0	6	0	**87**	0	0	0	7	0	0	0
OW	0	0	0	0	**92**	4	0	4	0	0	0
TW	0	0	0	0	9	**91**	0	0	0	0	0
BE	0	5	0	0	5	0	**84**	0	6	0	0
PJU	0	0	4	0	6	0	0	**86**	0	0	4
SMO	0	0	10	0	0	0	0	0	**90**	0	0
CL	0	0	0	0	0	0	0	0	0	**89**	11
BO	0	0	3	0	0	3	0	0	3	0	**91**
Average						**88.4**					

TABLE 3. The assessment results of the HAR system in the presence of graph description (while, in the absence of the proposed feature extraction and selection method) (Unit: %). Where WA is used for walking, RU is used for running, JU is used for jumping, SK is used for skipping, OW is used for one hand waving, TW is used for two hand waving, BE is used for bending, PJU is used for place jumping, SMO is used for side movement, CL is used for clapping, and BO is used for boxing.

Classes	WA	RU	JU	SK	OW	TW	BE	PJU	SMO	CL	BO
WA	**80**	5	0	4	0	6	4	0	0	1	0
RU	10	**84**	0	0	0	0	0	0	6	0	0
JU	0	5	**83**	0	7	0	0	5	0	0	0
SK	0	7	0	**86**	0	0	0	8	0	0	1
OW	0	0	0	0	**85**	0	5	0	10	0	0
TW	0	0	0	10	0	**90**	0	0	0	0	0
BE	0	0	8	0	0	0	**81**	0	7	0	4
PJU	0	0	5	0	0	7	0	**83**	0	0	5
SMO	0	0	9	0	3	0	0	0	**88**	0	0
CL	0	0	0	0	0	0	3	0	0	**89**	8
BO	0	0	0	0	0	0	0	0	0	8	**92**
Average						**85.5**					

IV. EMPLOYED DATASET

The proposed model was assessed against the dataset collected by kinect camera. Detailed description of the dataset is given below:

- *Depth Dataset using Kinect Camera:*
  This is a dataset of 550 video sequences collected (recorded) from overall 50 subjects. Each subject is a university student and performs various activities that include: running, walking, jumping, skipping, one hand waving, two hand waving, bending, place jumping, side movement, clapping, boxing. Every time, the dataset has been modified by incorporating newly activity frames in complex environments. The images of this dataset were from both male and female patients, and the age range of the patients were between 30 to 50 years. The original size of some activity frames are 320×280, and some are 480×320 pixel in others. Therefore, for the experiments, all the images in this dataset have been transformed to a vector with zero mean and the size 1×6400. Moreover, we reduce the

TABLE 4. The assessment results of the HAR system in the presence of local directional pattern (while, in the absence of the proposed feature extraction and selection method) (Unit: %). Where WA is used for walking, RU is used for running, JU is used for jumping, SK is used for skipping, OW is used for one hand waving, TW is used for two hand waving, BE is used for bending, PJU is used for place jumping, SMO is used for side movement, CL is used for clapping, and BO is used for boxing.

Classes	WA	RU	JU	SK	OW	TW	BE	PJU	SMO	CL	BO
WA	**91**	5	0	4	0	0	0	0	0	0	0
RU	0	**90**	0	0	0	6	0	4	0	0	0
JU	3	0	**88**	0	3	0	6	0	0	0	0
SK	0	7	0	**82**	0	0	0	11	0	0	0
OW	0	0	10	0	**87**	0	0	3	0	0	0
TW	0	0	7	0	0	**85**	0	0	8	0	0
BE	0	4	0	0	0	0	**86**	0	10	0	0
PJU	0	0	4	0	0	6	0	**83**	0	7	0
SMO	10	0	0	0	0	0	0	0	**90**	0	0
CL	0	2	0	2	0	0	4	0	0	**81**	11
BO	0	0	0	0	0	0	0	0	0	12	**88**
Average						**86.4**					

TABLE 5. The assessment results of the HAR system in the presence of projection histograms (while, in the absence of the proposed feature extraction and selection method) (Unit: %). Where WA is used for walking, RU is used for running, JU is used for jumping, SK is used for skipping, OW is used for one hand waving, TW is used for two hand waving, BE is used for bending, PJU is used for place jumping, SMO is used for side movement, CL is used for clapping, and BO is used for boxing.

Classes	WA	RU	JU	SK	OW	TW	BE	PJU	SMO	CL	BO
WA	**90**	0	0	0	0	6	0	4	0	0	0
RU	4	**88**	0	0	0	0	4	0	0	4	0
JU	0	5	**85**	0	0	7	0	0	3	0	0
SK	0	8	0	**82**	0	0	6	0	4	0	0
OW	6	0	11	0	**79**	0	0	4	0	0	0
TW	0	7	0	3	0	**82**	0	0	0	8	0
BE	0	0	10	0	9	0	**81**	0	0	0	0
PJU	0	0	0	0	0	6	0	**86**	0	8	0
SMO	7	0	4	0	0	0	0	0	**89**	0	0
CL	0	0	0	0	0	0	0	0	0	**90**	10
BO	0	0	0	0	0	0	0	0	0	8	**92**
Average						**85.8**					

TABLE 6. The assessment results of the HAR system in the presence of local directional pattern variance (while, in the absence of the proposed feature extraction and selection method) (Unit: %). Where WA is used for walking, RU is used for running, JU is used for jumping, SK is used for skipping, OW is used for one hand waving, TW is used for two hand waving, BE is used for bending, PJU is used for place jumping, SMO is used for side movement, CL is used for clapping, and BO is used for boxing.

Classes	WA	RU	JU	SK	OW	TW	BE	PJU	SMO	CL	BO
WA	**92**	8	0	0	0	0	0	0	0	0	0
RU	5	**93**	2	0	0	0	0	0	0	0	0
JU	5	0	**90**	0	0	5	0	0	0	0	0
SK	0	4	0	**91**	5	0	0	0	0	0	0
OW	0	0	6	0	**89**	0	0	5	0	0	0
TW	0	0	4	0	6	**86**	0	0	4	0	0
BE	0	0	0	0	0	5	**91**	0	0	4	0
PJU	0	0	0	0	3	0	0	**89**	0	8	0
SMO	0	7	0	0	0	0	0	0	**93**	0	0
CL	0	0	0	0	0	0	0	7	0	**87**	6
BO	0	0	0	0	0	0	0	0	0	8	**92**
Average						**90.3**					

size of each input frame to 80 × 80. In order to avoid unbalancing problem. The dataset was collected within the period of 4 months (from March to June of 2016).

V. EXPERIMENTAL SETUP

In this study, we performed the following set of experiments.

- In the first experiment, the performance of the proposed segmentation method is presented.
- The second experiment presents the accuracy of the proposed approach against the defined dataset collected by depth camera.
- This experiment presents number of sub-experiments. We used existing well-known feature extraction and selection techniques such as local binary pattern, graph description, local directional pattern, projection histograms, local directional pattern variance, contour

TABLE 7. The assessment results of the HAR system in the presence of contour profiles (while, in the absence of the proposed feature extraction and selection method) (Unit: %). Where WA is used for walking, RU is used for running, JU is used for jumping, SK is used for skipping, OW is used for one hand waving, TW is used for two hand waving, BE is used for bending, PJU is used for place jumping, SMO is used for side movement, CL is used for clapping, and BO is used for boxing.

Classes	WA	RU	JU	SK	OW	TW	BE	PJU	SMO	CL	BO
WA	**81**	9	0	4	0	4	0	2	0	0	0
RU	11	**82**	0	7	0	0	0	0	0	0	0
JU	0	4	**86**	6	0	4	0	0	0	0	0
SK	10	0	0	**90**	0	0	0	0	0	0	0
OW	0	3	0	5	**80**	0	7	0	5	0	0
TW	0	3	1	6	0	**83**	0	0	7	0	0
BE	0	6	0	4	0	0	**78**	8	0	0	4
PJU	1	6	0	7	0	4	0	**77**	0	5	0
SMO	0	0	0	0	0	0	0	0	**89**	4	7
CL	0	0	0	0	0	6	0	0	3	**81**	10
BO	0	0	0	0	5	0	0	4	3	9	**79**
Average						**82.4**					

TABLE 8. The assessment results of the HAR system in the presence of Hough transform (while, in the absence of the proposed feature extraction and selection method) (Unit: %). Where WA is used for walking, RU is used for running, JU is used for jumping, SK is used for skipping, OW is used for one hand waving, TW is used for two hand waving, BE is used for bending, PJU is used for place jumping, SMO is used for side movement, CL is used for clapping, and BO is used for boxing.

Classes	WA	RU	JU	SK	OW	TW	BE	PJU	SMO	CL	BO
WA	**76**	9	0	8	0	0	7	0	0	0	0
RU	10	**77**	0	5	0	5	0	0	3	0	0
JU	2	7	**80**	0	0	6	0	0	5	0	0
SK	0	9	0	**82**	0	4	0	0	5	0	0
OW	0	6	0	4	**75**	3	0	7	0	2	3
TW	5	0	3	0	6	**74**	0	8	0	4	0
BE	0	0	0	5	0	0	**78**	8	0	9	0
PJU	0	3	0	4	0	5	0	**79**	2	4	3
SMO	8	9	0	0	0	0	0	0	**83**	0	0
CL	0	0	0	0	0	0	0	6	0	**84**	10
BO	0	0	1	0	1	0	0	2	0	11	**85**
Average						**79.4**					

TABLE 9. The assessment results of the HAR system in the presence of geometric moment invariant (while, in the absence of the proposed feature extraction and selection method) (Unit: %). Where WA is used for walking, RU is used for running, JU is used for jumping, SK is used for skipping, OW is used for one hand waving, TW is used for two hand waving, BE is used for bending, PJU is used for place jumping, SMO is used for side movement, CL is used for clapping, and BO is used for boxing.

Classes	WA	RU	JU	SK	OW	TW	BE	PJU	SMO	CL	BO
WA	**87**	8	0	3	0	2	0	0	0	0	0
RU	10	**90**	0	0	0	0	0	0	0	0	0
JU	0	5	**91**	0	0	4	0	0	0	0	0
SK	10	5	0	**85**	0	0	0	0	0	0	0
OW	0	0	6	0	**89**	0	3	0	1	1	0
TW	1	3	0	4	0	**87**	2	0	0	0	3
BE	0	1	0	0	6	0	**92**	0	0	1	0
PJU	0	0	0	3	0	0	6	**91**	0	0	0
SMO	7	0	0	0	0	0	0	0	**93**	0	0
CL	0	1	0	2	0	0	4	0	1	**84**	8
BO	0	0	0	0	0	3	0	0	0	9	**88**
Average						**88.8**					

profiles, Hough transform, geometric moment invariant, wavelet transform, zernike moments, curvelet transform, scale invariant feature transform, spline curve approximation, template matching, Fourier descriptors, deformable templates, gradient feature, Gabor features, speed up robust feature, unitary image transforms, robust independent elementary features instead of using the proposed feature extraction method.

- In the last experiment, a set fo comparison results of the proposed system against previous works are presented. In this experiment, experimental code are borrowed for some works; while, some algorithms, we implemented with the exact settings described in their respective articles.

VI. RESULTS AND DISCUSSION

A. FIRST EXPERIMENT

As explained above, the performance of an HAR system is completely dependent on the segmentation of the human body, which is one of the most difficult tasks in literature.

TABLE 10. The assessment results of the HAR system in the presence of wavelet transform (while, in the absence of the proposed feature extraction and selection method) (Unit: %). Where WA is used for walking, RU is used for running, JU is used for jumping, SK is used for skipping, OW is used for one hand waving, TW is used for two hand waving, BE is used for bending, PJU is used for place jumping, SMO is used for side movement, CL is used for clapping, and BO is used for boxing.

Activities	WA	RU	JU	SK	OW	TW	BE	PJU	SMO	CL	BO
WA	**90**	6	2	0	2	0	0	0	0	0	0
RU	6	**91**	0	0	0	2	0	0	3	0	0
JU	1	0	**89**	4	5	0	1	0	0	0	0
SK	0	1	7	**92**	0	0	0	0	0	0	0
OW	0	3	0	2	**93**	0	0	0	2	0	0
TW	0	0	3	0	0	**94**	0	0	0	3	0
BE	0	0	3	0	4	0	**88**	0	0	0	5
PJU	0	2	0	0	0	5	0	**90**	0	3	0
SMO	6	1	0	0	0	0	0	0	**87**	6	0
CL	0	0	0	0	0	0	0	0	0	**94**	6
BO	0	0	1	0	3	0	2	0	0	4	**90**
Average						**90.73**					

TABLE 11. The assessment results of the HAR system in the presence of zernike moments (while, in the absence of the proposed feature extraction and selection method) (Unit: %). Where WA is used for walking, RU is used for running, JU is used for jumping, SK is used for skipping, OW is used for one hand waving, TW is used for two hand waving, BE is used for bending, PJU is used for place jumping, SMO is used for side movement, CL is used for clapping, and BO is used for boxing.

Activities	WA	RU	JU	SK	OW	TW	BE	PJU	SMO	CL	BO
WA	**82**	8	0	0	4	0	3	0	3	0	0
RU	10	**80**	0	5	0	0	0	5	0	0	0
JU	4	1	**79**	0	6	0	4	0	2	3	1
SK	0	3	0	**84**	0	6	0	4	0	0	3
OW	0	0	9	0	**88**	0	3	0	0	0	0
TW	0	0	0	6	0	**87**	0	0	7	0	0
BE	0	0	0	4	0	0	**78**	11	0	2	5
PJU	0	2	3	0	0	10	0	**77**	0	8	0
SMO	0	0	11	0	0	4	0	0	**85**	0	0
CL	0	0	3	0	0	0	0	0	0	**89**	8
BO	0	0	0	0	0	0	6	0	0	10	**84**
Average						**83.00**					

TABLE 12. The assessment results of the HAR system in the presence of curvelet transform (while, in the absence of the proposed feature extraction and selection method) (Unit: %). Where WA is used for walking, RU is used for running, JU is used for jumping, SK is used for skipping, OW is used for one hand waving, TW is used for two hand waving, BE is used for bending, PJU is used for place jumping, SMO is used for side movement, CL is used for clapping, and BO is used for boxing.

Activities	WA	RU	JU	SK	OW	TW	BE	PJU	SMO	CL	BO
WA	**93**	7	0	0	0	0	0	0	0	0	0
RU	5	**94**	1	0	0	0	0	0	0	0	0
JU	0	0	**96**	2	0	0	2	0	0	0	0
SK	2	0	5	**89**	0	0	0	0	2	0	2
OW	0	0	5	0	**88**	0	3	0	0	4	0
TW	0	0	0	0	6	**94**	0	0	0	0	0
BE	0	0	0	3	0	0	**87**	7	0	3	0
PJU	0	0	0	0	0	0	0	**90**	0	5	5
SMO	4	2	0	0	2	0	0	0	**92**	0	0
CL	0	0	0	0	0	0	0	0	0	**93**	7
BO	0	0	0	0	0	0	3	0	0	6	**91**
Average						**91.55**					

For this reason, we recommend an unsupervised and powerful AC model, which combines two energy functions that automatically detect and segment the human body. With human activity videos, the evolution of active contour in a frame is accomplished independently of the supplementary frames, which means that in a video, the human body segmenting is done through frame-based. The only information in the final contour from the previous activity frame is used to determine the starting position of the active contour in the present activity frame. Originally, an ellipse with major axis along y-axis of length 45 and minor axis along x-axis of length 65 is selected as the initial contour. For the entire experiments, the initial shape is same for the whole frames but only the center positions vary. In every video, the initial image is separated heuristically, means that the initial contour is nearer to the human body.

TABLE 13. The assessment results of the HAR system in the presence of scale invariant feature transform (while, in the absence of the proposed feature extraction and selection method) (Unit: %). Where WA is used for walking, RU is used for running, JU is used for jumping, SK is used for skipping, OW is used for one hand waving, TW is used for two hand waving, BE is used for bending, PJU is used for place jumping, SMO is used for side movement, CL is used for clapping, and BO is used for boxing.

Activities	WA	RU	JU	SK	OW	TW	BE	PJU	SMO	CL	BO
WA	**88**	10	0	0	0	1	0	0	1	0	0
RU	11	**89**	0	0	0	0	0	0	0	0	0
JU	0	4	**84**	0	5	0	0	3	0	4	0
SK	2	6	0	**92**	0	0	0	0	0	0	0
OW	0	0	0	0	**91**	0	0	0	4	0	5
TW	0	0	0	0	4	**93**	0	3	0	0	0
BE	0	0	0	0	0	0	**89**	11	0	0	0
PJU	0	1	0	3	0	0	6	**90**	0	0	0
SMO	0	5	0	0	0	0	0	0	**95**	0	0
CL	0	0	0	0	0	0	0	4	0	**84**	12
BO	0	0	0	0	0	0	0	0	0	15	**85**
Average						**89.05**					

TABLE 14. The assessment results of the HAR system in the presence of spline curve approximation (while, in the absence of the proposed feature extraction and selection method) (Unit: %). Where WA is used for walking, RU is used for running, JU is used for jumping, SK is used for skipping, OW is used for one hand waving, TW is used for two hand waving, BE is used for bending, PJU is used for place jumping, SMO is used for side movement, CL is used for clapping, and BO is used for boxing.

Activities	WA	RU	JU	SK	OW	TW	BE	PJU	SMO	CL	BO
WA	**92**	2	4	0	1	0	1	0	0	0	0
RU	3	**93**	0	2	0	1	0	0	1	0	0
JU	0	3	**97**	0	0	0	0	0	0	0	0
SK	0	10	0	**90**	0	0	0	0	0	0	0
OW	0	0	0	0	**89**	9	0	2	0	0	0
TW	0	0	0	0	7	**93**	0	0	0	0	0
BE	0	2	0	4	0	0	**88**	5	0	1	0
PJU	0	0	0	2	4	0	3	**85**	0	6	0
SMO	0	4	0	0	2	0	0	0	**94**	0	0
CL	0	0	0	0	0	0	0	0	0	**92**	8
BO	0	0	0	0	0	0	2	0	0	7	**91**
Average						**91.27**					

TABLE 15. The assessment results of the HAR system in the presence of template matching (while, in the absence of the proposed feature extraction and selection method) (Unit: %). Where WA is used for walking, RU is used for running, JU is used for jumping, SK is used for skipping, OW is used for one hand waving, TW is used for two hand waving, BE is used for bending, PJU is used for place jumping, SMO is used for side movement, CL is used for clapping, and BO is used for boxing.

Activities	WA	RU	JU	SK	OW	TW	BE	PJU	SMO	CL	BO
WA	**78**	10	0	2	3	2	1	0	4	0	0
RU	13	**77**	5	0	1	0	1	2	0	1	0
JU	4	5	**75**	15	0	0	0	0	1	0	0
SK	0	7	0	**81**	0	0	6	0	0	5	1
OW	0	0	0	0	**82**	0	0	0	0	0	0
TW	0	0	0	0	14	**83**	0	3	0	0	0
BE	0	4	0	1	1	4	**76**	5	3	2	4
PJU	0	4	0	2	0	3	0	**82**	0	3	6
SMO	7	9	0	4	0	0	0	0	**80**	0	0
CL	0	0	0	0	2	1	0	2	0	**87**	8
BO	0	0	0	3	0	2	2	0	3	13	**77**
Average						**79.82**					

Then from the second frame, the center location of the initial contour in the present frame is the average value of the points along with the final contour in the former frame. Consider that along the final contour of frame $j (j \geq 1)$, there are N points $\left(y_i^{(j)}, z_i^{(j)}\right)$, $i = 1..N$. Then, the center $(k_y^{(j+1)}$, $k_z^{(j+1)})$ of the initial contour in the frame $(j+1)$is calculated as

$$\binom{(j+1)}{\underset{y}{k}} = \frac{1}{N}\sum_{i=1}^{N} \overset{(j)}{\underset{i}{y}};\ \binom{(j+1)}{\underset{z}{k}} = \frac{1}{N}\sum_{i=1}^{N} \overset{(j)}{\underset{i}{z}}$$

Some sample segmentation results of the proposed segmentation model along with the above-mentioned rule on four different types of dynamic activities such as skipping, walking, running, and jumping are presented in Figure 3-B1. Furthermore, for fair comparison, we presented the sample segmentation results of the existing algorithm such as (Chan-Vese) in Figure 5.

As can be seen from Figures 3-B1 and 5 that the proposed segmentation model showed best performance against existing model (Chan-Vese active model).

TABLE 16. **The assessment results of the HAR system in the presence of Fourier descriptors (while, in the absence of the proposed feature extraction and selection method) (Unit: %). Where WA is used for walking, RU is used for running, JU is used for jumping, SK is used for skipping, OW is used for one hand waving, TW is used for two hand waving, BE is used for bending, PJU is used for place jumping, SMO is used for side movement, CL is used for clapping, and BO is used for boxing.**

Activities	WA	RU	JU	SK	OW	TW	BE	PJU	SMO	CL	BO
WA	**90**	3	4	1	1	0	1	0	0	0	0
RU	5	**92**	0	0	1	1	0	0	1	0	0
JU	0	6	**89**	5	0	0	0	0	0	0	0
SK	0	5	7	**88**	0	0	0	0	0	0	0
OW	0	0	0	0	**93**	7	0	0	0	0	0
TW	0	0	0	0	3	**94**	2	0	1	0	0
BE	0	0	0	0	0	0	**91**	9	0	0	0
PJU	0	0	0	2	0	0	6	**92**	0	0	0
SMO	0	5	0	0	0	0	0	0	**95**	0	0
CL	0	0	0	0	0	0	0	0	0	**87**	13
BO	0	0	0	0	1	1	3	0	0	9	**86**
Average						**90.64**					

TABLE 17. **The assessment results of the HAR system in the presence of deformable templates (while, in the absence of the proposed feature extraction and selection method) (Unit: %). Where WA is used for walking, RU is used for running, JU is used for jumping, SK is used for skipping, OW is used for one hand waving, TW is used for two hand waving, BE is used for bending, PJU is used for place jumping, SMO is used for side movement, CL is used for clapping, and BO is used for boxing.**

Activities	WA	RU	JU	SK	OW	TW	BE	PJU	SMO	CL	BO
WA	**75**	10	10	2	1	1	0	1	0	0	0
RU	9	**79**	7	5	0	0	0	0	0	0	0
JU	0	7	**82**	5	2	0	1	1	1	0	1
SK	3	5	0	**85**	0	4	0	3	0	0	0
OW	0	0	3	0	**83**	9	0	0	5	0	0
TW	0	0	6	0	7	**81**	2	1	0	2	1
BE	0	0	0	0	0	0	**88**	0	0	7	5
PJU	0	1	0	0	0	0	6	**84**	0	3	6
SMO	0	10	0	0	0	0	0	0	**90**	0	0
CL	0	0	0	0	0	0	0	0	0	**77**	23
BO	0	0	0	0	0	0	4	2	0	10	**84**
Average						**82.54**					

TABLE 18. **The assessment results of the HAR system in the presence of gradient feature (while, in the absence of the proposed feature extraction and selection method) (Unit: %). Where WA is used for walking, RU is used for running, JU is used for jumping, SK is used for skipping, OW is used for one hand waving, TW is used for two hand waving, BE is used for bending, PJU is used for place jumping, SMO is used for side movement, CL is used for clapping, and BO is used for boxing.**

Activities	WA	RU	JU	SK	OW	TW	BE	PJU	SMO	CL	BO
WA	**79**	11	4	6	0	0	0	0	0	0	0
RU	8	**77**	3	6	1	0	1	1	0	1	2
JU	2	5	**82**	10	0	0	0	0	0	0	0
SK	1	5	3	**86**	0	1	1	2	0	1	0
OW	3	0	4	0	**80**	6	1	1	2	3	0
TW	0	0	1	0	12	**78**	0	0	4	2	3
BE	0	0	0	0	0	0	**76**	8	0	7	9
PJU	1	0	0	2	3	0	3	**83**	2	6	0
SMO	3	4	0	5	0	1	1	0	**85**	1	0
CL	0	0	3	2	1	0	1	2	3	**74**	14
BO	0	1	2	1	1	0	3	1	1	17	**73**
Average						**79.36**					

B. SECOND EXPERIMENT

The first experiment focuses on evaluating and obtaining the results of MEMM model on the input dataset. This is necessary to stress the significance of the model, and to evaluate against other models over naturalistic depth datasets of eleven different activities. The entire results are presented in Table 1 and Figure 5.

It can be seen in Table 1 and Figure 5 that the proposed model showed best performance against naturalistic dataset (collected by depth camera) that has eleven different types of activities.

C. THIRD EXPERIMENT

In this evaluation, we present and show the importance of the proposed feature extraction and selection method in a common HAR system. For this purpose, we utilized well-known existing feature extraction and selection methods in the absence of the developed algorithm. The entire evaluation results are described in Tables 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, and 22 respectively.

As can be seen in Tables 2 through 22 that in the absence of the proposed feature extraction and selection method, the HAR system did not achieve significant classification

TABLE 19. **The assessment results of the HAR system in the presence of Gabor features (while, in the absence of the proposed feature extraction and selection method) (Unit: %). Where WA is used for walking, RU is used for running, JU is used for jumping, SK is used for skipping, OW is used for one hand waving, TW is used for two hand waving, BE is used for bending, PJU is used for place jumping, SMO is used for side movement, CL is used for clapping, and BO is used for boxing.**

Activities	WA	RU	JU	SK	OW	TW	BE	PJU	SMO	CL	BO
WA	**91**	9	0	0	0	0	0	0	0	0	0
RU	4	**88**	3	2	0	0	0	3	0	0	0
JU	3	2	**93**	0	0	0	2	0	0	0	0
SK	3	1	4	**88**	1	2	0	0	1	0	0
OW	0	0	0	0	**92**	4	0	0	0	2	2
TW	0	0	0	0	5	**90**	1	0	1	2	1
BE	0	0	0	0	0	0	**95**	0	0	3	2
PJU	0	0	0	0	0	0	0	**87**	0	13	0
SMO	2	5	3	4	0	0	0	0	**86**	0	0
CL	0	0	0	0	0	0	0	0	0	**94**	6
BO	0	0	0	0	0	0	0	0	0	4	**96**
Average						**90.91**					

TABLE 20. **The assessment results of the HAR system in the presence of speed up robust feature (while, in the absence of the proposed feature extraction and selection method) (Unit: %). Where WA is used for walking, RU is used for running, JU is used for jumping, SK is used for skipping, OW is used for one hand waving, TW is used for two hand waving, BE is used for bending, PJU is used for place jumping, SMO is used for side movement, CL is used for clapping, and BO is used for boxing.**

Activities	WA	RU	JU	SK	OW	TW	BE	PJU	SMO	CL	BO
WA	**91**	3	4	2	0	0	0	0	0	0	0
RU	3	**97**	0	0	0	0	0	0	0	0	0
JU	2	1	**94**	1	0	1		0	1	0	0
SK	0	7	0	**93**	0	0	0	0	0	0	0
OW	0	0	0	0	**95**	5	0	0	0	0	0
TW	0	0	0	0	0	**90**	0	0	0	7	3
BE	0	0	0	0	1	0	**92**	4	0	1	2
PJU	0	1	0	1	0	1	2	**89**	0	4	2
SMO	0	6	4	2	0	0	0	0	**88**	0	0
CL	0	0	0	0	0	1	0	1	0	**93**	5
BO	0	0	0	0	0	0	0	0	0	4	**96**
Average						**92.54**					

TABLE 21. **The assessment results of the HAR system in the presence of unitary image transforms (while, in the absence of the proposed feature extraction and selection method) (Unit: %). Where WA is used for walking, RU is used for running, JU is used for jumping, SK is used for skipping, OW is used for one hand waving, TW is used for two hand waving, BE is used for bending, PJU is used for place jumping, SMO is used for side movement, CL is used for clapping, and BO is used for boxing.**

Activities	WA	RU	JU	SK	OW	TW	BE	PJU	SMO	CL	BO
WA	**85**	7	4	3	0	1	0	0	0	0	0
RU	10	**82**	3	0	2	1	0	1	1	0	0
JU	2	3	**84**	4	0	2	1	0	1	2	1
SK	1	1	3	**88**	2	0	0	0	5	0	0
OW	2	3	1	4	**80**	2	1	2	1	3	1
TW	0	0	0	0	13	**87**	0	0	0	0	0
BE	0	1	0	1	0	3	**90**	1	0	2	2
PJU	1	1	1	3	1	2	1	**82**	0	3	5
SMO	5	3	1	2	1	0	4	0	**81**	2	1
CL	0	0	0	0	0	0	0	0	0	**89**	11
BO	0	0	0	0	1	0	2	1	3	7	**86**
Average						**84.91**					

rates. Because most of the existing methods suffer from various issues (as described in Section II). On the other hand, the proposed HAR system showed significant performance in the presence of the proposed feature extraction and selection method. This is because the proposed approach has the ability to extract and select the informative features from the activity frames by taking the advantages of the forward selection and backward regression models. For such purpose, we selected 300 features as a feature vector and fed it to the classifier.

D. FOURTH EXPERIMENT

In the last experiment, we compared the classification rate of the proposed HAR system against the existing works [12], [34], [52]–[58]. For some of the works, we implemented their respective works based on the setting as mentioned in Section V. While, for some systems, we borrowed their implementations for fair comparison. The entire assessments of the proposed approach along with the previous studies are represented in Table 23.

TABLE 22. The assessment results of the HAR system in the presence of robust independent elementary features (while, in the absence of the proposed feature extraction and selection method) (Unit: %). Where WA is used for walking, RU is used for running, JU is used for jumping, SK is used for skipping, OW is used for one hand waving, TW is used for two hand waving, BE is used for bending, PJU is used for place jumping, SMO is used for side movement, CL is used for clapping, and BO is used for boxing.

Activities	WA	RU	JU	SK	OW	TW	BE	PJU	SMO	CL	BO
WA	**90**	4	2	1	0	2	0	1	0	0	0
RU	7	**89**	0	4	0	0	0	0	0	0	0
JU	0	5	**85**	2	3	0	1	4	0	0	0
SK	0	5	0	**91**	0	0	0	0	4	0	0
OW	0	0	0	0	**92**	4	0	0	0	1	3
TW	0	0	0	0	4	**93**	0	0	0	2	1
BE	2	1	2	0	3	1	**86**	1	2	0	2
PJU	0	0	3	2	1	0	2	**88**	0	1	3
SMO	3	0	2	3	0	1	0	0	**91**	0	0
CL	0	0	0	0	0	0	0	0	0	**93**	7
BO	0	0	0	0	0	4	0	0	0	7	**89**
Average						**89.73**					

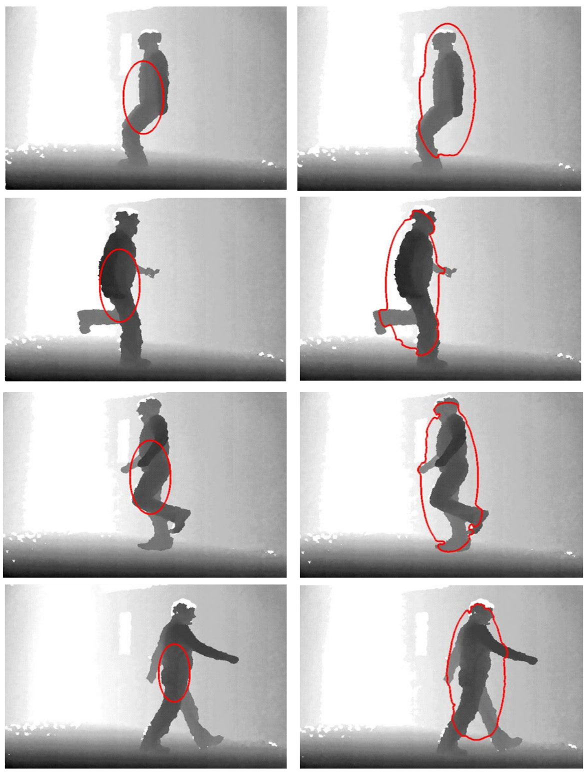

FIGURE 4. Segmentation Results of the CV AC model on dynamic activities (like first row is jumping activity (with $\beta = 1.0$, $\gamma = 0.1$, $\eta = 0.0$, and cpu time = 54s), second row is running activity (with $\beta = 1.0$, $\gamma = 0.1$, $\eta = 0.0$, and cpu time = 65s), third row is skipping activity(with $\beta = 1.0$, $\gamma = 0.1$, $\eta = 0.0$, and cpu time = 45s), and the fourth row is walking activity (with $\beta = 1.0$, $\gamma = 0.1$, $\eta = 0.0$, and cpu time = 85s) respectively). Similarly, the first column is the initial contour; while, the second column is final contour.

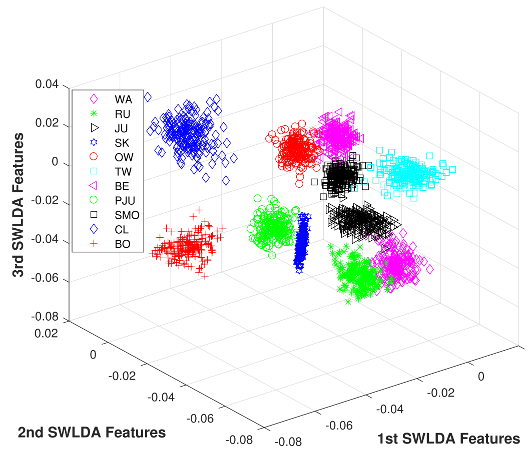

FIGURE 5. Classification of the proposed model in 3D sample space. Where WA is used for walking, RU is used for running, JU is used for jumping, SK is used for skipping, OW is used for one hand waving, TW is used for two hand waving, BE is used for bending, PJU is used for place jumping, SMO is used for side movement, CL is used for clapping, and BO is used for boxing.

As noted from Table 23 that the developed approach in the human activity recognition system achieved utmost accuracy compared to the existing works under the depth dataset.

E. DISCUSSION

Activity recognition is a key element in defining a user's context for delivering services based on the kind of ubiquitous tools. Accordingly, this study presents an accurate and efficient activity recognition system that has the ability to accurately recognize human activities using depth camera against naturalistic environment. There are two methods that have been proposed for the system. One method is for segmentation module; while, the other one is for feature extraction module.

For the segmentation, we proposed a vigorous human body segmentation algorithm, which is an unsupervised method that efficiently perceives and segments the human body from the video frame. The proposed segmentation algorithm is the association of two energy functions (Chan-Vese and Battacharya functions) which not only reduces the similarities within the human body parts but also increases the detachment between the human body and the background. The proposed segmentation technique shows utmost performance on static and dynamic activities, which is one of the limitations of the previous works.

TABLE 23. The assessment results of developed system together with the existing works on the depth dataset collected by depth camera (Unit: %).

Existing Works	Classification Rates	Standard Deviation
[52]	89.7	±2.2
[12]	92.1	±1.9
[53]	93	±0.8
[54]	88.6	±3.5
[55]	87.4	±4.1
[56]	90.4	±1.4
[57]	86.0	±4.8
[34]	91.9	±2.9
[58]	88.1	±3.1
Developed System	**97.3**	±2.7

Similarly, for the feature extraction, the use of a new robust feature extraction technique named Stepwise Linear Discriminant Analysis (SWLDA) has been proposed. The stepwise regression method starts with choosing an equation containing the single best X variable and then attempts to build more X's one at a time as long as these conditions are worthwhile. The order of addition and selection is based on partial $F-$test values in order to know which variable should enter first or next. Then the highest partial F-value is compared to the default value.

After the addition process or forward entering the deletion process or backward deletion starts. In this process partial test value for all predictor variable already in the Queue were calculated. Then the lowest partial test value is compared to pre-selected or default significance level, i.e., if the lower value if less than the corresponding value, then remove that variable, and then start again the process of calculating F-Test, otherwise adopt the regression equation.

VII. CONCLUSION AND FUTURE DIRECTION

Currently, in healthcare environment, the privacy is a significant concern with an application to real time activity recognition systems. In most of the previous researches, 2D cameras were used to capture the human activities that may compromise the privacy of the data. Therefore, depth camera is considered a good candidate for this kind of issues.

As a result, we have made two important contributions to this work. First contribution in segmentation module, which depends on association of two energy functions (Chan-vese and Bhattacharyya functions) coupled with level set method. (1) Chan-Vese function is utilized to reduce the variations within the human body, (2) Bhattacharyya distance function is employed in order to enlarge the distance amongst human body and the background. First, the original contour must be closer to the human body. The proposed active model uses this initial contour to identify and remove the human body from the current image, and the final contour of this image is transferred to the next image. The diameter of the transferred contour is used as the center of the original contour in the next frame and then repeats the process until the last image in the activity video. As a result, the proposed AC approach used the designed scheme that can accurately identify the human body from the activity video. There are three parameters such as γ, β and η employed in the proposed technique. Γ smooths the contour. η moves the contour in a standard direction, and β weights the restraints of within-body similarity and between-body variations.

The second contribution is for feature extraction and selection. For this purpose, we proposed a non-linear feature extraction and selection technique which has the ability to choose the finest informative features with the help of using the forward selection algorithm. Furthermore, the proposed technique also removes the inappropriate features with the help of using backward regression method. The proposed approach starts from only finest X variable and then efforts to increase further X_s one by one, if the conditions are adequate. The selection and removal depend upon the partial $F-$test value to find which variable should be passed in. The uppermost partial $F-$value is compared to the default F-to-enter value. After this, forward selection or backward removal starts. In this step, partial test values for entire variables previously in the queue are estimated. Then the least partial test value is compared to the pre-selected consequence levels like F_0, (means, if $F_L < F_0$, then the variable Z_L will be eliminated, and the F-test will start again; otherwise, the regression equation is adopted).

We have trained and tested the proposed approach offline. In the future research, we will implement the proposed approach in healthcare domain (like in hospitals) to tackle the privacy and security issues.

ACKNOWLEDGMENT

The authors would like to extend their sincere appreciation to the Central Laboratory at Jouf University to support this study.

REFERENCES

[1] R. Saini, P. Kumar, P. P. Roy, and D. P. Dogra, "A novel framework of continuous human-activity recognition using Kinect," *Neurocomputing*, vol. 311, pp. 99–111, Oct. 2018.

[2] R. Saini, P. Kumar, B. Kaur, P. P. Roy, D. P. Dogra, and K. C. Santosh, "Kinect sensor-based interaction monitoring system using the BLSTM neural network in healthcare," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 9, pp. 2529–2540, Sep. 2019.

[3] S. Mukherjee, R. Saini, P. Kumar, P. P. Roy, D. P. Dogra, and B.-G. Kim, "Fight detection in hockey videos using deep network," *J. Multimedia Inf. Syst.*, vol. 4, no. 4, pp. 225–232, 2017.

[4] N. P. Tiilikainen, *A Comparative Study of Active Contour Snakes*. Copenhagen, Denmark: Copenhagen Univ., 2007, pp. 21–26.

[5] M. Z. Uddin and T.-S. Kim, "Continuous hidden Markov models for depth map-based human activity recognition," in *Hidden Markov Models, Theory and Applications*, 2011, pp. 225–247.

[6] M. Z. Uddin, J. Lee, and T.-S. Kim, "Shape-based human activity recognition using independent component analysis and hidden Markov model," in *Proc. Int. Conf. Ind., Eng. Other Appl. Appl. Intell. Syst.*, Springer, 2008, pp. 245–254.
[7] M. H. Siddiqi, M. Fahim, S. Lee, and Y.-K. Lee, "Human activity recognition based on morphological dilation followed by watershed transformation method," in *Proc. Int. Conf. Electron. Inf. Eng.*, Aug. 2010, pp. V2–433.
[8] M. Z. Uddin, J. J. Lee, and T.-S. Kim, "Independent shape component-based human activity recognition via hidden Markov model," *Int. J. Speech Technol.*, vol. 33, no. 2, pp. 193–206, Oct. 2010.
[9] M. Zia Uddin, T.-S. Kim, and J. Tai Kim, "Video-based indoor human gait recognition using depth imaging and hidden Markov model: A smart system for smart home," *Indoor Built Environ.*, vol. 20, no. 1, pp. 120–128, Feb. 2011.
[10] A. Jalal, M. Z. Uddin, J. T. Kim, and T.-S. Kim, "Recognition of human home activities via depth silhouettes and $\Re$ transformation for smart homes," *Indoor Built Environ.*, vol. 21, no. 1, pp. 184–190, Feb. 2012.
[11] R. Ramli, N. Zakaria, and P. Sumari, "Privacy issues in pervasive healthcare monitoring system: A review," *World Acad. Sci. Eng. Technol*, vol. 72, no. 12, pp. 741–747, 2010.
[12] 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, p. 2886, May 2020.
[13] S. Kamal and A. Jalal, "A hybrid feature extraction approach for human detection, tracking and activity recognition using depth sensors," *Arabian J. Sci. Eng.*, vol. 41, no. 3, pp. 1043–1051, Mar. 2016.
[14] A. Jalal, Y. Kim, S. Kamal, A. Farooq, and D. Kim, "Human daily activity recognition with joints plus body features representation using Kinect sensor," in *Proc. Int. Conf. Informat., Electron. Vis. (ICIEV)*, Jun. 2015, pp. 1–6.
[15] W. Ding, B. Hu, H. Liu, X. Wang, and X. Huang, "Human posture recognition based on multiple features and rule learning," *Int. J. Mach. Learn. Cybern.*, vol. 11, no. 11, pp. 2529–2540, Nov. 2020.
[16] C. Tang, H. Hu, W. Wang, W. Li, H. Peng, and X. Wang, "Using a multilearner to fuse multimodal features for human action recognition," *Math. Problems Eng.*, vol. 2020, pp. 1–18, Aug. 2020.
[17] N. Jaouedi, F. J. Perales, J. M. Buades, N. Boujnah, and M. S. Bouhlel, "Prediction of human activities based on a new structure of skeleton features and deep learning model," *Sensors*, vol. 20, no. 17, p. 4944, Sep. 2020.
[18] M. Siddiqi, R. Ali, M. Rana, E.-K. Hong, E. Kim, and S. Lee, "Video-based human activity recognition using multilevel wavelet decomposition and stepwise linear discriminant analysis," *Sensors*, vol. 14, no. 4, pp. 6370–6392, Apr. 2014.
[19] M. H. Siddiqi, A. M. Khan, and S.-W. Lee, "Active contours level set based still human body segmentation from depth images for video-based activity recognition," *KSII Trans. Internet Inf. Syst.*, vol. 7, no. 11, pp. 2839–2852, 2013.
[20] L. Xia, C.-C. Chen, and J. K. Aggarwal, "Human detection using depth information by Kinect," in *Proc. CVPR Workshops*, Jun. 2011, pp. 15–22.
[21] T. Fabry, D. Smeets, and D. Vandermeulen, "Surface representations for 3D face recognition," in *Face Recognition*, Apr. 2010, pp. 273–294.
[22] L. Cinque, A. Danani, P. Dondi, and L. Lombardi, "Real-time foreground segmentation with Kinect sensor," in *Int. Conf. Image Anal. Process.*, Springer, 2015, pp. 56–65.
[23] A. Abramov, K. Pauwels, J. Papon, F. Worgotter, and B. Dellen, "Depth-supported real-time video segmentation with the Kinect," in *Proc. IEEE Workshop Appl. Comput. Vis. (WACV)*, Jan. 2012, pp. 457–464.
[24] P. K. Wang, C. H. Chen, S. H. Pun, B. Zhang, P. U. Mak, M. I. Vai, and T. C. Lei, "Parallel architecture to accelerate superparamagnetic clustering algorithm," *Electron. Lett.*, vol. 56, no. 14, pp. 701–704, Jul. 2020.
[25] R. Saini, P. Pratim Roy, and D. Prosad Dogra, "A segmental HMM based trajectory classification using genetic algorithm," *Expert Syst. Appl.*, vol. 93, pp. 169–181, Mar. 2018.
[26] J.-H. Kim, B.-G. Kim, P. P. Roy, and D.-M. Jeong, "Efficient facial expression recognition algorithm based on hierarchical deep neural network structure," *IEEE Access*, vol. 7, pp. 41273–41285, 2019.
[27] H. Lee, S. Lee, and O. Choi, "Improved method on image stitching based on optical flow algorithm," *Int. J. Eng. Bus. Manage.*, vol. 12, Jan. 2020, Art. no. 184797902098092.
[28] Z. Yang, L. Leng, and W. Min, "Extreme downsampling and joint feature for coding-based palmprint recognition," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021.
[29] J. Chu, Z. Guo, and L. Leng, "Object detection based on multi-layer convolution feature fusion and online hard example mining," *IEEE Access*, vol. 6, pp. 19959–19967, 2018.
[30] L. Leng, M. Li, C. Kim, and X. Bi, "Dual-source discrimination power analysis for multi-instance contactless palmprint recognition," *Multimedia Tools Appl.*, vol. 76, no. 1, pp. 333–354, Jan. 2017.
[31] L. Leng, S. Zhang, X. Bi, and M. K. Khan, "Two-dimensional cancelable biometric scheme," in *Proc. Int. Conf. Wavelet Anal. Pattern Recognit.*, Jul. 2012, pp. 164–169.
[32] L. Leng and J. Zhang, "Palmhash code for palmprint verification and protection," in *Proc. 25th IEEE Can. Conf. Electr. Comput. Eng. (CCECE)*, Apr. 2012, pp. 1–4.
[33] M. R. Keyvanpour, N. Khanbani, and Z. Aliniya, "Detection of individual activities in video sequences based on fast interference discovery and semi-supervised method," *Multimedia Tools Appl.*, vol. 80, pp. 13879–13910, Apr. 2021.
[34] C. Chen, R. Jafari, and N. Kehtarnavaz, "A real-time human action recognition system using depth and inertial sensor fusion," *IEEE Sensors J.*, vol. 16, no. 3, pp. 773–781, Feb. 2016.
[35] L. He, X. Yang, L. Lu, W. Wu, A. Ahmad, and G. Jeon, "A novel multi-focus image fusion method for improving imaging systems by using cascade-forest model," *EURASIP J. Image Video Process.*, vol. 2020, no. 1, p. 5, Dec. 2020.
[36] Y. Liu, X. Chen, H. Peng, and Z. Wang, "Multi-focus image fusion with a deep convolutional neural network," *Inf. Fusion*, vol. 36, pp. 191–207, Jul. 2017.
[37] N. Ahmed, J. I. Rafiq, and M. R. Islam, "Enhanced human activity recognition based on smartphone sensor data using hybrid feature selection model," *Sensors*, vol. 20, no. 1, p. 317, Jan. 2020.
[38] N. C. Silva, J. M. Moreira, and P. Menezes, "Features selection for human activity recognition with iPhone inertial sensors," in *Proc. Portuguese Conf. Artif. Intell.*, 2013, pp. 1–11.
[39] S. D. Jadhav and H. P. Channe, "Comparative study of K-NN, naive Bayes and decision tree classification techniques," *Int. J. Sci. Res.*, vol. 5, no. 1, pp. 1842–1845, 2016.
[40] J. Han, M. Kamber, and J. Pei, "Data mining concepts and techniques third edition," *Morgan Kaufmann Ser. Data Manage. Syst.*, vol. 5, no. 4, pp. 83–124, 2011.
[41] Y. Tian, J. Zhang, J. Wang, Y. Geng, and X. Wang, "Robust human activity recognition using single accelerometer via wavelet energy spectrum features and ensemble feature selection," *Syst. Sci. Control Eng.*, vol. 8, no. 1, pp. 83–96, Jan. 2020.
[42] N. Hoque, M. Singh, and D. K. Bhattacharyya, "EFS-MI: An ensemble feature selection method for classification," *Complex Intell. Syst.*, vol. 4, no. 2, pp. 105–118, Jun. 2018.
[43] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. R. Mullers, "Fisher discriminant analysis with kernels," in *Proc. Neural Netw. Signal Process. 9th. IEEE Signal Process. Soc. Workshop*, Aug. 1999, pp. 41–48.
[44] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Comput.*, vol. 12, no. 10, pp. 2385–2404, Oct. 2000.
[45] S. Mika, "Kernel Fisher discriminants," Tech. Rep., 2003.
[46] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.
[47] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. 1, no. 4, pp. 321–331, Jan. 1988.
[48] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.
[49] D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "Toward enhanced P300 speller performance," *J. Neurosci. Methods*, vol. 167, no. 1, pp. 15–21, Jan. 2008.
[50] L. E. Baum, "An equality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes," *Inequalities*, vol. 3, pp. 1–8, Jan. 1972.
[51] F. Samaria, "Face recognition using hidden Markov models," Ph.D. dissertation, Univ. Cambridge, Cambridge, U.K., 1994.
[52] M. Gavrilova, Y. Wang, F. Ahmed, and P. Paul, "Kinect sensor gesture and activity recognition for consumer cognitive systems," *IEEE Consum. Electron. Mag., Special Issue Consum. Electron.*, vol. 4, pp. 88–96, 2017.
[53] D. He and L. Li, "A new Kinect-based posture recognition method in physical sports training based on urban data," *Wireless Commun. Mobile Comput.*, vol. 2020, pp. 1–9, Apr. 2020.

[54] H. Wang, C. Dartigues-Pallez, and M. Riveill, "Supervised learning for human action recognition from multiple Kinects," in *Proc. Int. Conf. Database Syst. Adv. Appl.*, Springer, 2020, pp. 32–46.
[55] C. M. Ranieri, S. MacLeod, M. Dragone, P. A. Vargas, and R. A. F. Romero, "Activity recognition for ambient assisted living with videos, inertial units and ambient sensors," *Sensors*, vol. 21, no. 3, p. 768, Jan. 2021.
[56] S. Liu and H. Wang, "Action recognition using key-frame features of depth sequence and ELM," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 10, pp. 52–56, 2017.
[57] B. Ghojogh, H. Mohammadzade, and M. Mokari, "Fisherposes for human action recognition using Kinect sensor data," *IEEE Sensors J.*, vol. 18, no. 4, pp. 1612–1627, Feb. 2018.
[58] K. Kim, A. Jalal, and M. Mahmood, "Vision-based human activity recognition system using depth silhouettes: A smart home system for monitoring the residents," *J. Electr. Eng. Technol.*, vol. 14, no. 6, pp. 2567–2573, Nov. 2019.

MUHAMMAD HAMEED SIDDIQI received the Bachelor of Computer Science degree (Hons.) from the Islamia College (Chartered University), Peshawar, Pakistan, in 2007, and the master's and Ph.D. degrees from the Ubiquitous Computing (UC) Laboratory, Department of Computer Engineering, Kyung Hee University, Suwon, South Korea, in 2012 and 2016, respectively. From 2008 to 2009, he was a Graduate Assistant with the Universiti Teknologi PETRONAS, Malaysia. From March 2016 to August 2016, he was also a Postdoctoral Research Scientist with the Department of Computer Science and Engineering, Sungkyunkwan University, Suwon, South Korea. From September 2016 to October 2020, he was an Assistant Professor with the Department Computer Science, Jouf University, Sakaka, Saudi Arabia. Since November 2020, he has been working as an Associate Professor with the Department Computer Science, Jouf University. He published more than 70 research articles in highly reputable international journals and conferences. His research interests include image processing, pattern recognition, machine intelligence, activity recognition, and facial expression recognition. He is also a reviewer for different journals and conferences.

NABIL ALMASHFI received the B.S. degree in computer and information sciences from Jouf University, Saudi Arabia, in 2009, the M.S. degree in computer science from Saint Joseph's University, USA, in 2013, and the Ph.D. degree in computer science and informatics from Oakland University, USA, in 2020. He is currently an Assistant Professor with the College of Computer and Information Sciences, Jouf University. Prior to his recent appointment as an Assistant Professor, he was a Lecturer with Jouf University. His research interests include software engineering and artificial intelligence.

AMJAD ALI received the B.S. and M.S. degrees in computer science from the COMSATS Institute of Information Technology, Pakistan, in 2006 and 2008, respectively, and the Ph.D. degree from the Electronics and Radio Engineering Department, Kyung Hee University, South Korea, in 2015. From 2018 to 2019, he was a Postdoctoral Research Scientist with the UWB Wireless Communications Research Center (formerly the Key National IT Research Center), Department of Information and Communication Engineering, Inha University, South Korea, and also with the Mobile Network and Communications Laboratory, School of Electrical Engineering, Korea University, Seoul, South Korea. Since 2015, he has been an Assistant Professor with the Department of Computer Science, COMSATS University Islamabad, Lahore Campus, Pakistan. He has published many peer-reviewed international journal articles and conference papers. His main research interests include the Internet of Things, cognitive radio networks (interference modeling, dynamic spectrum access, power and admission control, spectrum management, spectrum trading, MAC protocols, performance modeling, and optimization), 5G cellular networks (spectrum/resource management, coexistence, distributed wireless access, scheduling, power control, network selection, and mobility/handover management), *ad hoc* social networks multimedia cloud computing, smart grid, and vehicular networks. He received the Excellent Research Contribution Award from the UWB Wireless Communications Research Center, for his excellent research performances. He is serving as a Reviewer for *IEEE Communications Magazine*, IEEE Transactions on Industrial Informatics, IEEE Transactions on Circuit and Systems for Video Technology, IEEE Transactions on Vehicular Technology, IEEE Journal on Selected Areas in Communications, IEEE Internet of Things Journal, IEEE Access, *IET Communications*, and *Computer Networks*. He is also serving as an Editorial Board Member for IEEE Access.

MADALLAH ALRUWAILI received the bachelor's degree (Hons.) from Jouf University, Saudi Arabia, in 2005, the M.S. degree from the University of Science, Malaysia, in 2009, and the Ph.D. degree from Southern Illinois University, Carbondale, IL, USA, in 2015. His Ph.D. dissertation entitled "Enhancement and Restoration of Dust Images." He is currently an Assistant Professor of computer engineering and networks with Jouf University. He is also the Dean of the College of Computer and Information Sciences. His research interests include image processing, image quality analysis, pattern recognition, computer vision, and biomedical imaging.

YOUSEF ALHWAITI received the bachelor's degree (Hons.) in computer science from Jouf University, Saudi Arabia, in 2010, the master's degree in computer science from Ball State University, USA, in 2012, and the Ph.D. degree from Pace University, USA, in 2019. He is currently an Assistant Professor of computer science. He is also working with Jouf University. He has published many articles, most of them in machine learning. His research interests include deep learning, pattern recognition, and computer vision.

SAAD ALANAZI received the B.S. degree in computer science from Jouf University, Saudi Arabia, in 2007, the M.S. degree in computer science from Ball State University, in 2011, and the Ph.D. degree in artificial intelligence from Staffordshire University, in 2017. He is currently an Assistant Professor with the Department of Computer Science, Jouf University. His research interests include natural language processing, text mining, and image processing.

M. M. KAMRUZZAMAN (Senior Member, IEEE) received the B.E. and M.S degrees in computer science and engineering and the Ph.D. degree in information and communication technology. He worked as a Postdoctoral Research Fellow with Shenzhen University, China. He is currently working with Jouf University, Saudi Arabia. His research interests include 5G, artificial intelligence, image processing, remote sensing, GIS, cloud computing, and big data. He is serving as a member of editorial board of several international journals. He is also serving as a TPC and a reviewer of few international journals and conferences.