IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 18, 2021, accepted December 1, 2021, date of publication December 3, 2021, date of current version December 10, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3132559

Maximum Entropy Markov Model for Human Activity Recognition Using Depth Camera

IBRAHIM ALRASHDI, MUHAMMAD HAMEED SIDDIQI, YOUSEF ALHWAITI, MADALLAH ALRUWAILI, AND MOHAMMAD AZAD
College of Computer and Information Sciences, Jouf University, Sakaka, Al-Jouf 72388, Saudi Arabia

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

This work was supported by the Deanship of Scientific Research at Jouf University under Grant DSR 2020-06-3670.

ABSTRACT Activity recognition is an essential factor in the determination of daily routine of a human being. There exist numerous Human Activity Recognition (HAR) systems; however, an HAR with a practical accuracy is still in search and a challenge at large. The classifiers utilized in existing systems offer degrading recognition rates in various key environments such as depth camera environment among others. To address the limitation of degrading recognition rates, in this paper, we propose Maximum Entropy Markov Model (MEMM) that solves the degrading recognition rate problem. In MEMM, we model the states of activity recognition as the states of the model itself i.e. as the states of MEMM, and hence consider the observations of video-sensor as the observations of MEMM. Further, we use a modified version of Viterbi, a machine learning algorithm, to generate the most likely probable state sequence based on these observations. Then, from such a state sequence, we use MEMM to predict the activity state. We evaluate the performance MEMM against depth dataset having eleven different types of activities in a large-scale experimentation process. The results show that MEMM outperforms existing well-known methods by achieving a weighted average recognition rate 96.3% across the naturalistic dataset collected using depth camera.

INDEX TERMS Healthcare, machine learning, activity recognition, depth camera, MEMM, state-of-the-art.

I. INTRODUCTION

Remote activity recognition and analysis has advanced and brought fruitful additions and benefits to the development in telemedicine and e-health domains. These e-services and tools assist clinicians in their diagnosis and decision making process by monitoring daily routines (activities, exercises etc.) of their patients remotely, and have proven helpful in diseases such as stroke. Tele-stroke in [1], presents the case of monitoring acute stroke patients using telemedicine. Usually, the states of stroke patients can be determined using their activities by applying efficient activity recognition techniques, and daily exercises can be recommended. For example, walking an running can be identified helping them and then regular remote care can help them better manage their routine. Moreover, a psychiatrist can use the telemedicine technology for the treatment of a patient with post-traumatic stress disorder (PTSD) by monitoring his/her activities remotely [2]. Telemedicine has not only been studied for cases related strokes, stress etc.; however, it also useful for medical specialist where they use activity recognition to heart patients and their failures [3], hence a prime candidate for activities analysis. Among these different application areas, boost the stamina need more research in telemedicine to efficiently monitor patients and get most of the technology at hand today. In summary, video-based activity recognition is an observable indication of person's sentimental state, mental activity and behavior [4].

Privacy of the collected data, either personal information or activity information, is of great concern to the patients or the activity performers. The concerns by the subjects are just and they include sensitivity to the possibility of sharing of information with or without consent and may be exposed to threats. A study in [5] goes in depth in these concerns. In this research, we use depth-cameras because, unlike RGB-cameras, depth cameras do not reveal the identity of the subject or its other sensitive information. To the best of our knowledge, we know of limited research in activity recognition using the depth-cameras.

A typical activity recognition system consists of four basic modules: pre-processing module, feature extraction, feature selection, and feature recognition modules. The

The associate editor coordinating the review of this manuscript and approving it for publication was Farid Boussaid.

pre-processing module diminishes the environmental factors such as lighting and luminance. Then, the feature extraction module extracts distinguishing features from each activity shape and quantizes it as a discrete symbol. The feature selection module selects a subset of relevant features from a large number of features extracted in the feature extraction module. Finally, in the recognition module, first a classifier is trained with the training data and then generate appropriate labels for the activity frames in the incoming video data (as in production) as shown in Figure 1.

Although, plenty of research contribution exist in the area of HAR focusing on improving the feature extraction and feature selection stages [6]–[15], most of these HAR systems utilize well-known conventional learning methods such as artificial neural networks (ANN), support vector machine (SVM), hidden Markov model (HMM), deep learning, hidden conditional random fields (HCRF), etc. Several research studies suggest HMM to be used method and comparatively perform better as compared to others, see [8]–[10], [13] for details. However, in some other research studies such as speech recognition [16], gesture recognition [17], [18] show that HMM, a generative learning model, is not as accurate as expected because of its Markovian property. The Markovian property presumes that the state (current) only depends on the previous state [13].

Motivated by the limitation of the learning model above, we have proposed the use of MEMM for the HAR problem. Therefore, in this research, we present the design of a more accurate and robust HAR system to recognize human activities. We utilized the methods that were proposed in some of our previous works for feature extraction and selection, called "wavelet transform coupled with optical flow and stepwise linear discriminant analysis (SWLDA)" [19]. Just as for classification, we propose a new recognition model where the activity states are modeled as the states of the maximum entropy Markov model (MEMM). Similarly, in this model, we consider the video-sensor observations as the observations of MEMM. We use a modified Viterbi, a machine-learning algorithm, to produce the most probable activity state sequence based on these observations. Moreover, from the most likely state sequence, we predict the activity state is predicted through our stated algorithm. We evaluate and validate our proposed approach on standard action datasets which were collected using Microsoft depth camera in controlled environment. Our proposed approach outperforms the existing state-of-the-art systems/approaches by having a higher average recognition rate of 96.3% across the datasets. It is important to state here that, this study is the first to utilize MEMM model as a classifier for HAR systems.

The rest of the paper is organized as: Section II summarized state-of-the-art activity recognition system using depth camera. Section III presented 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

In literature, there is a huge amount of systems have been developed for video-based human activity recognition that only focused on feature extraction and feature selection modules [13], [20]–[24]. Most of these systems employed the most utilized classifiers including artificial neural networks, support vector machine, Gaussian mixture model, hidden Markov model, deep learning. Among them, hidden Markov model is one of the strong candidates for classification.

Hidden Markov model is used for impulsive-based classification. In this classifier, the feature-level information is utilized in order to handle the series data and that is one the main advantages of HMM. Apart from this, some other classifiers do not have this property which might help them in learning the sequence of the feature vectors and these models are named as vector-based models. However, HMM is using common property due to which it assumes that the previous state is dependent on present state, and because of this the two connecting states labels theoretically happen successively in the final sequence that is not all the time satisfied in reality [25]. Moreover, HMM are reproductive in environment that directed it towards the individuality expectations within the states and annotations [26].

On the contrary, authors in [27] proposed a system for video human activity recognition, and they classified the activities by employing Naive Bayes classifier. The system has been trained and tested on five publicly available standards datasets. However, the Naive Bayes classifier treats its variables independent from each other. One could resolve this limitation by first performing some sort of statistical analysis to find possible correlations between variables/features. And then, choose only the variables that have the least correlation [28].

Similarly, the human activities have been recognized in video surveillance by employing K-Means, Fuzzy C-Means, Multilayer Perceptron Self-Organizing Maps and Feed-forward Neural networks. They claimed better accuracy in various domains [29]. However, However, this approach has some common limitations such as easiness to fall under resident minimum, the static learning rate, difficulty to find the number of neurons in the hidden layer [30]. Furthermore, in [31], using the background subtraction approach, HOG features and Back-Propagation Neural Network (BPNN) classifier, they were able to recognize human activities. Moreover, it uses the mean filter to obtain the background of the image and the areas of the image containing important information. Furthermore, to extract features, it uses the histogram of oriented gradients (HOG) [32] descriptor coupled with local shape information and intensity gradients or edge directions. Finally, for classification, they employed BPNN. However, BPNN has two major limitations, being the low convergence rate and the instability [33]. These limitations are the results of being trapped in a local minimum [34], [35]

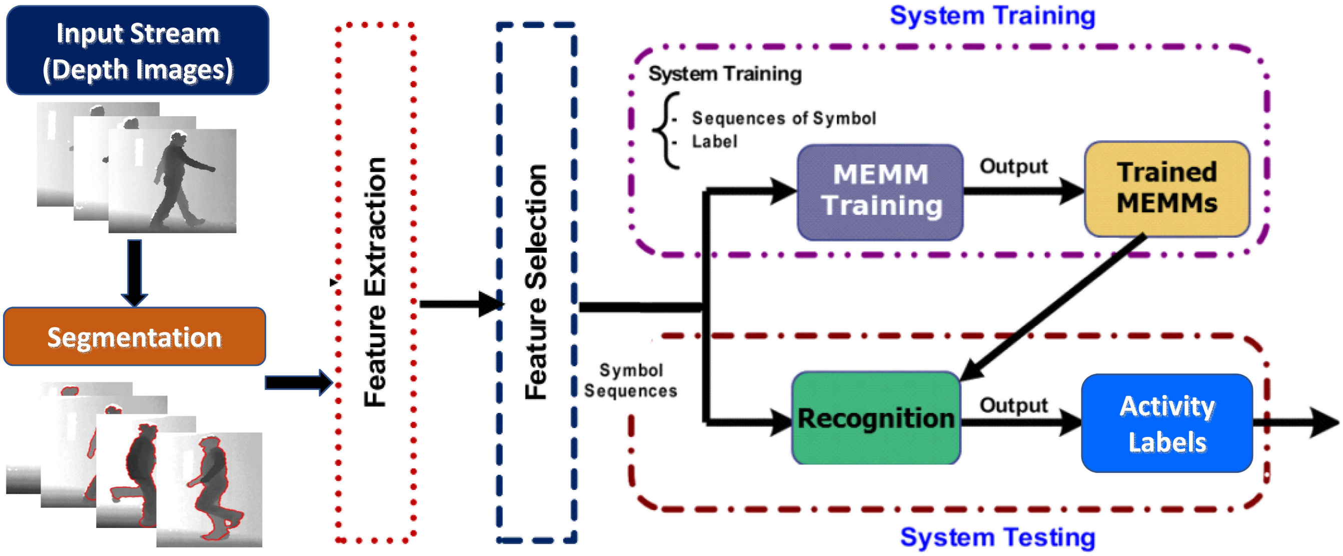

FIGURE 1. Typical human activity recognition system using depth camera.

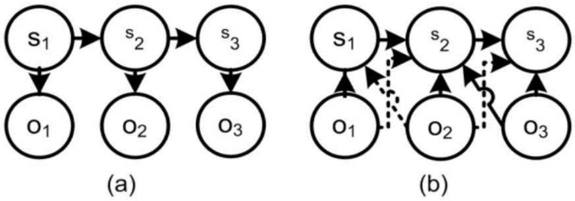

FIGURE 2. (a) Presents the dependence graph of HMM, whereas (b) shows the dependence graph of MEMM [41].

and the possibility in overshooting the minimum of the error surface [33].

Another system was proposed by [15] in order to recognize the human activities in video surveillance. In their system, in order to extract features (as a set), they employed thresholding-based method coupled with inverse Haar wavelet transform. After that, they used the K-nearest neighbors (kNN) algorithm to recognize the activity from the given input data. However, kNN has two major problems when it comes to implementation:

- It is a lazy learning method and
- It depends on the selection of the value of k [36].

Other limitations present in this method corresponds to the high memory consumption which limits its application [37].

In summary, there lots of existing works have been presented to get significant performance by utilizing the existing well-known models; however, everyone has its own limitations which didn't address until yet.

In this paper, we have proposed the MEMM model that will solve most of the limitations of the existing classifiers. In particular, the classifiers that are most widely used in HAR systems and the feature selection, extraction related to HAR systems.

III. MATERIALS AND METHODS

As discussed before, a human activity recognition (HAR) system consists of four basic modules. For the first three modules (such as pre-processing, feature extraction and selection), we utilized the methods from our previous works. While, for the recognition module, we proposed a new recognition model. Each of them are described as below.

A. HUMAN BODY SEGMENTATION

In the first module, we employed one of our previous unsupervised segmentation technique [38] in order to segment the human body from the video frame against depth camera. This approach was the combination of two energy functions like Chan-Vese and Bhattacharya functions that eliminates the dissimilarities among the human body and enlarge the distance between the human body and the background respectively.

B. FEATURE EXTRACTION AND SELECTION

We utilized wavelet transform to extract the prominent and key features [19] (like movable features) from different parts of the body. To do so, we employ the symlet wavelet transform coupled with optical flow. This approach helps in diminishing noise before extraction of features for activity movement.

Though the above approach may result in extraction of good and required features; however, there is still the possibility of redundancy among those extracted features. To address this redundancy, we apply the stepwise linear discriminant analysis (SWLDA) [39] to the extracted feature space. SWLDA is able to choose the most informative features as it takes advantage of forward selection model, and it can also remove the irrelevant features using the backward regression model.

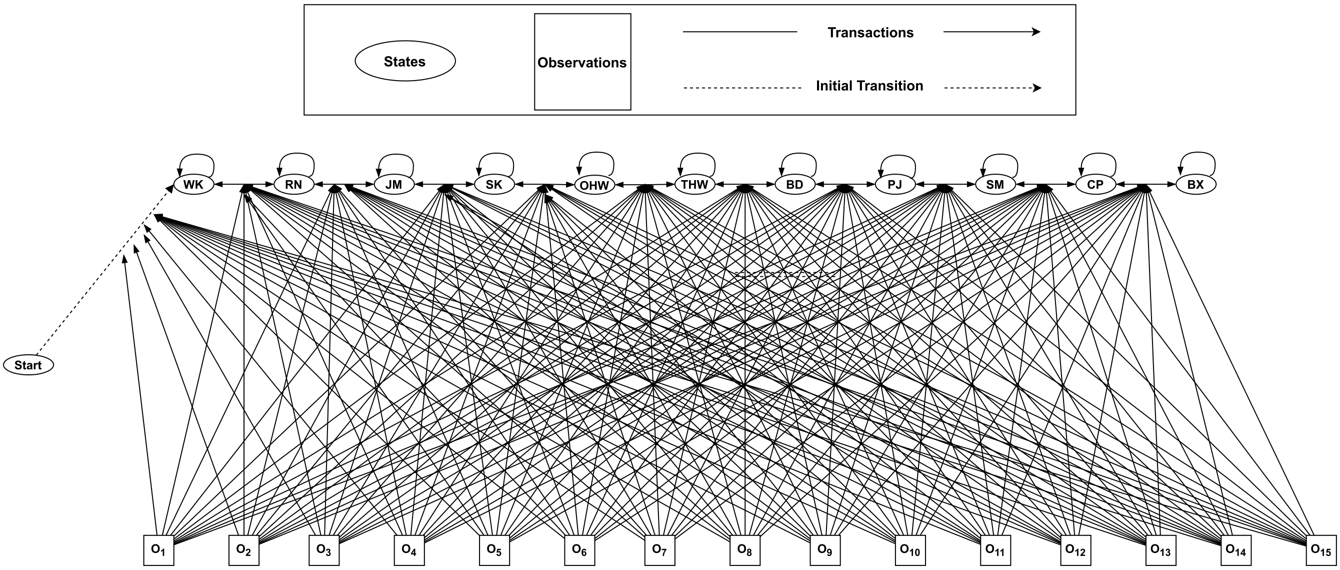

FIGURE 3. Activity state model based on MEMM for activity recognition system. Where WK for walking, RN for running, JM for jumping, SK for skipping, OHW for one hand waving, THW for two hand waving, BD for bending, PJ for place jumping, SM for side movement, CP for clapping, and BX for boxing.

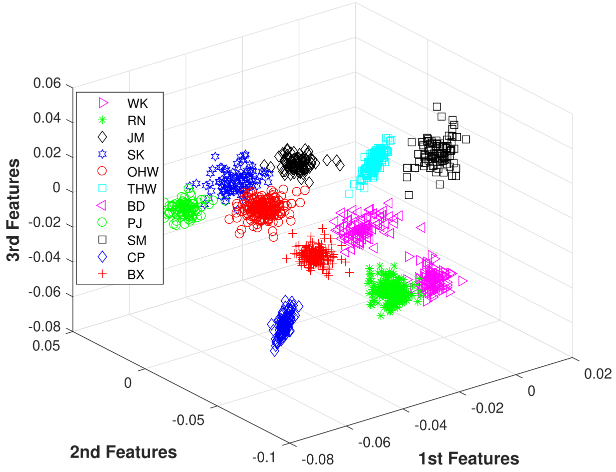

FIGURE 4. Classification of the proposed model in 3D sample space. Where WK for walking, RN for running, JM for jumping, SK for skipping, OHW for one hand waving, THW for two hand waving, BD for bending, PJ for place jumping, SM for side movement, CP for clapping, and BX for boxing.

C. PROPOSED MAXIMUM ENTROPY MARKOV MODEL (MEMM)

The proposed model consists of several steps. We present these steps one by one in the following.

1) PROCEDURE OF MEMM

In this paper, we model the activity states as Maximum Entropy Markov Model (MEMM). To the best of our knowledge, it is considered as one of the reasonable and essential alternative compared to Hidden Markov Model (HMM) for simulating the ordered states and/or observations. The difference between these two approaches is how to compute the maximum probability of the likelihood of the sequence of the observance. Three joint probabilities are used in the case of generative HMM. Only one conditional probability is used for the case of MEMM [40]. We can easily see the difference of HMM and MEMM from the dependency graph (shown in Figure 2) shows the dependencies between the states and observations in MEMM and HMM models.

The M states of MEMM model can be seen in Figure 3. The notations are demonstrated below:

- The set of states $\{E_1, E_2, \ldots, E_p\}$: It can be represented as the human activities $E = \{E_1, E_2, \ldots, E_p\}$ = walking; running; jumping, skipping, one hand waving, two hand waving, bending, place jumping, side movement, clapping, boxing, etc.
- The frame observations $\{O_1, O_2, \ldots, O_T\}$: It can be represented as $O = \{O_1, O_2, \ldots, O_T\}$, where T is the duration of the observation.
- The set of features at each observation O_i is the vector of observed features $\{f_1, f_2, \ldots, f_n\}$, which are extracted from the human activity frames at the time slot t.
- The total number of features is N.

The main objective here is to determine $S_L = \{S_1, S_2, \ldots, S_k\}$ being the highly likely state sequence where $S_i \in E$ based on O, where O is the current sequence of observations O over the duration T. To generate S_L, HMM requires transition, emission, and initial probabilities as $P(Si|Si - 1)$, $P(O_i|S_i)$, and $P(S_i)$ respectively. Whereas, MEMM requires only the probability $P(S_i|S_{i-1}, O_i)$, which can be obtained easily from the maximum entropy model as we will discuss in the following section. This is the reason of why we use this MEMM model.

2) THE PROCEDURE OF PARAMETER ASSESSMENT IN MEMM

Even though there are many techniques are available in the theory for the assessment of MEMM parameters (e.g., described in details [40]), this paper uses the technique of the maximum entropy model (1). This model ($MaxEnt: H$) is used to assess the probability of the transformation between states S_{i-1} and S_i based on the observation O.

$$P\left(S_i|S_{i-1}, O_i\right) = \frac{e^{\sum_{k=1}^{N} w_k f_k}}{R} \tag{1}$$

In 1, we represent the total count of features by N, the weights of the logistic regression by w and the feature value of the training data by f. Using the rules of probability axiom (i.e. the sum of the probabilities of the whole state space must equals to 1), we can normalize the right hand side of the equation 1 by R, and subsequently, we can get the value of R from equations 2 and 3).

$$P\left(S_i|S_{i-1}, O_i\right) = \frac{e^{\sum_{k=1}^{N} w_k f_k}}{\sum_i P\left(S_i|S_{i-1}, O_i\right)} \tag{2}$$

$$P\left(S_i|S_{i-1}, O_i\right) = \frac{e^{\sum_{k=1}^{N} w_k f_k}}{\sum_i e^{\sum_{k=0}^{N-1} w_k f_k}} \tag{3}$$

According to 3, the $MaxEnt(H)$ parameter w_k is now the major concern to find out $P(S_i|S_(i-1), O_i)$. This is mainly because the parameter f_k (feature parameter) is known from the training dataset. Since we use the activity classes/labels as the states of our model i.e. MEMM model, to define the activity's label, the probability of classes/labels should be greater than the other labels. Hence, we formalize the maximization of $P(S_i|S_(i-1), O_i)$ using parameter w as optimization problem as shown in the following 4.

$$\hat{w} = \underset{w}{argmax} \frac{e^{\sum_{k=1}^{N} w_k f_k}}{\sum_i e^{\sum_{k=1}^{N} w_k f_k}} \tag{4}$$

Algorithm 1: Assessment of MEMM Parameters (S, O)

```
Begin
Initialize $T \leftarrow S = \{E_1, E_2, \ldots, E_p\}$
Randomly select a state $E_i$
While $T$ do
Find all pairs of state-observation $(E_i, O_i)$ from training
    dataset.
Consider the selected $E_i$ as the state $S_{i-1}$ in determining
    $P(S_i|S_{i-1}, O_i)$
Determine optimal weight parameter $w$ from 7 through
    L-BFGS optimization method to maximize the log
    likelihood probability $P(S_i|S_{i-1}, O_i)$
$T \leftarrow T - E_i$
Select a state $E_i$ from $T$
End while
End
```

Algorithm 2: Adapted Viterbi (H, S, O)

```
Begin
$p = |s|$
$i = 1$
While $(i \leq p)$
  $V_1(i) \ = P(S_i|O_1)$
  $D_1(i) \ = 0$
  $i = i + 1$
End while
$t = 2$
While $(t \leq T)$
  $j = 1$
    While $(j \leq p)$
      $V_t(j) = \underset{1 \leq k \leq p}{max} \left(V_{t-1} \times (k) \ \times P(E_j|E_k, O_t)\right)$
      $D_t(j) = \underset{1 \leq k \leq p}{argmax} \left(V_{t-1} \times (k) \ \times P(E_j|E_k, O_t)\right)$
      $j = j + 1$
    End while
    $t = t + 1$
End while
$j^* = \underset{1 \leq j \leq p}{max} \ (V_T(j))$
$t_T = j_T^* = \underset{1 \leq j \leq M}{argmax} \ (V_T(j))$
$t = T - 1$
While $(t \geq 1)$
  $i_t^* = \underset{\tau+1}{D} \ (j_{\tau+1}^*)$
  $t_\tau = i_\tau^*$
  $t = t - 1$
End while
Return $S_L$
End
```

Let M be the total instances in the training dataset and consider the log likelihood probability, we can write equation 4 as 5 as follows:

$$\hat{w} = \underset{w}{argmax} \sum_{j}^{M} log \frac{e^{\sum_{k=1}^{N} w_k f_k^j}}{\sum_i e^{\sum_{k=0}^{N-1} w_k f_k^j}} \tag{5}$$

Next, we use the regularization to penalize the large value of w.

$$\hat{w} = \underset{w}{argmax} \sum_{j}^{M} log \frac{e^{\sum_{k=1}^{N} w_k f_k^j}}{\sum_i e^{\sum_{k=1}^{N} w_k f_k^j}} - \beta \times R(w) \tag{6}$$

In the above equation, we use the Gaussian distribution $N(\mu, \sigma^2)$ of parameter w for regularization as shown in 7.

$$\hat{w} = \underset{w}{argmax} \sum_{j}^{M} log \frac{e^{\sum_{k=1}^{N} w_k f_k^j}}{\sum_i e^{\sum_{k=1}^{N} w_k f_k^j}} - \sum_{j}^{M} \frac{(w_k - \mu_k)^2}{2\sigma_k^2} \tag{7}$$

Since we have obtained equation 7 to be a log-sum exponential equation, we employ the widely used and popular method called Broyden Fletcher Goldfarb Shanno (BFGS) to learn the optimal weight parameter w of MEMM. We choose BFGS because it supports unconstrained optimization. We describe the training process in algorithm 1.

3) THE PROCEDURE TO GENERATE THE HUMAN ACTIVITY STATE SEQUENCES USING VITERBI ALGORITHM

We start by using the algorithm Viterbi (as shown in Algorithm 2 to determine the hidden humanity state sequence that is more plausible. We do so from a given tuple of observations O. At each instant of time t_i, we extract features from the video frame at t_i and consider that frame as an observation O_i. We use (8) to determine the Viterbi value V.

$$V_t(j) = \underset{1\leq k\leq p}{max} \left(V_{t-1} \times (k)\ \times P(E_j|E_k, O_t)\right) \tag{8}$$

Here, state j is in $1 \leq j \leq p$. However, we make use of optimal weight parameter w from training system to determine $P(E_j|E_k, O_t)$ 3. In regards to the observation O, our modified Viterbi algorithm gives the set of most likely states for activities as $S_L = \{S_1, S_2, \ldots, S_k\}$ where each $S_i \in E$.

At the end, we infer the predicted human activity sequence. This is done by inferring the highly likely activity state sequence S over the duration T.

4) THE PROCEDURE OF THE CLASSIFICATION OF HUMAN ACTIVITY STATES

Below we describe the Algorithm 3 that explains the procedure to classify human activity sequences from generated activity states sequence. T duration is used to vary several video frames. Moreover, we use cardinalities for each state and they are determined to define the states of activities over the overall duration T. We measure the distinct states cardinality i.e., $|E_1|, |E_2|, \ldots, |E_p|$ from S and the highest cardinality activity state is defined as the predicted activity.

IV. UTILIZED DATASET

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

- *Depth Dataset Using Kinect Camera:*
  This is a dataset of 550 video sequences collected (recorded) from overall 50 subjects under Microsoft kinect camera. Each subject is a university student and performs various activities that include: running, walking, jumping, skipping, one and 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 subjects 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 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).

Algorithm 3: Classification Human Activity (H, S, O, R)

```
Begin
$S_L =$ Viterbi $(H, S, O)$
$p = |S|$
$i = 1$
While $(i \leq p)$
  $F_{E_i} = 0$
  $Q = |S_L|$
  $j = 1$
  While $(j \leq Q)$
    if $(E_i == S_j)$
      $F_{E_i} = F_{E_i} + 1$
    End if
  End while
  $|E_i| = F_{E_i}$
End while
$\hat{E} = \underset{E_i}{argmax}|E_i|$
$i = 1$
While $(i \leq p)$
  If $(E_i \in \hat{E})$
    If $(|E_i| > R_1$ && $\hat{E} \in \{‘Walking'\})$
      return $\hat{E}$
    Else If $(|E_i| > R_2$ && $\hat{E} \in \{‘Running'\})$
      return $\hat{E}$
    Else If $(|E_i| > R_3$ && $\hat{E} \in \{‘Jumping'\})$
      return $\hat{E}$
    Else If $(|E_i| > R_4$ && $\hat{E} \in \{‘Skipping'\})$
      return $\hat{E}$
    Else If $(|E_i| > R_5$ && $\hat{E} \in \{$‘One Hand Waving’$\})$
      return $\hat{E}$
    Else If $(|E_i| > R_6$ && $\hat{E} \in \{$‘Two Hand Waving’$\})$
      return $\hat{E}$
    Else If $(|E_i| > R_7$ && $\hat{E} \in \{‘Bending'\})$
      return $\hat{E}$
    Else If $(|E_i| > R_8$ && $\hat{E} \in \{$‘Place Jumping’$\})$
      return $\hat{E}$
    Else If $(|E_i| > R_9$ && $\hat{E} \in \{$‘Side Movement’$\})$
      return $\hat{E}$
    Else If $(|E_i| > R_10$ && $\hat{E} \in \{$‘Clapping’$\})$
      return $\hat{E}$
    Else If $(|E_i| > R_11$ && $\hat{E} \in \{$‘Boxing’$\})$
      return $\hat{E}$
    End if
  End if
Else return $argmax_{E_i}|E_i|$
End while
End
```

V. EXPERIMENTAL SETUP

In this study, we evaluate our model on the results of extensive experimentations in order to stress the importance and

TABLE 1. Recognition rates of the proposed model (Unit: %). Where WK for walking, RN for running, JM for jumping, SK for skipping, OHW for one hand waving, THW for two hand waving, BD for bending, PJ for place jumping, SM for side movement, CP for clapping, and BX for boxing.

Activities	WK	RN	JM	SK	OHW	THW	BD	PJ	SM	CP	BX
WK	**95**	2	1	1	0	0	0	0	1	0	0
RN	1	**94**	1	1	0	0	1	0	2	0	0
JM	1	1	**97**	0	1	0	0	0	0	0	0
SK	2	1	0	**94**	0	0	1	0	2	0	0
OHW	0	0	0	0	**98**	2	0	0	0	0	0
THW	0	0	0	0	2	**97**	1	0	0	0	0
BD	0	0	0	0	0	0	**99**	1	0	0	0
PJ	0	0	0	0	0	0	0	**100**	0	0	0
SM	2	1	1	0	1	0	0	0	**95**	0	0
CP	0	0	0	0	0	0	1	2	0	**94**	3
BX	0	0	0	0	0	1	0	1	0	2	**96**
Average						**96.3**					

TABLE 2. Recognition rates of the system against k-nearest neighbor (kNN) (having k = 4 value) instead of using the proposed model (Unit: %). Where WK for walking, RN for running, JM for jumping, SK for skipping, OHW for one hand waving, THW for two hand waving, BD for bending, PJ for place jumping, SM for side movement, CP for clapping, and BX for boxing.

Activities	WK	RN	JM	SK	OHW	THW	BD	PJ	SM	CP	BX
WK	**59**	6	4	5	4	3	3	4	3	4	5
RN	5	**62**	3	2	4	4	2	5	6	2	5
JM	3	4	**67**	4	3	5	2	3	2	4	3
SK	4	5	3	**61**	5	3	3	5	6	2	3
OHW	2	4	5	2	**69**	3	4	2	3	2	2
THW	5	6	5	5	6	**58**	3	3	5	2	2
BD	3	5	2	5	4	3	**66**	2	4	1	5
PJ	6	2	2	6	3	5	2	**64**	4	2	4
SM	4	4	6	2	5	3	2	6	**60**	7	1
CP	3	3	3	3	3	3	3	3	3	**70**	3
BX	4	3	5	5	3	1	7	2	3	5	**62**
Average						**63.4**					

reliability of our model for HAR systems. All the experiments are performed on a PC with configuration Intel® Pentium® Core^{TM} i7-6700 (3.4 GHz) having a RAM capacity of 16 GB for the entire experiments.

- In this experiment, we evaluate our model on a 10−fold cross-validation scheme. In total, we have 10 subjects. In which, we use one subject's data as test data whereas the data of remaining 9 subjects is used as the training dataset. We perform this experiment 10 times and each time the test data is different i.e. the test subject is chosen different each time.
- In this second experiment, we repeat the entire experiments in the absence of our model over existing model. i.e., we evaluate the following existing machine learning models: Random Forest, K-nearest neighbors Support Vector Machine, Decision Trees, Artificial Neural Network, Naive Baise, Logistic Regression, Markov Chain, Gaussian Mixture Model, Hidden Conditional Random Field, and Recurrent Neural Networks.
- In the end, we compare the accuracies of the existing systems as opposed to the results of our proposed model.

VI. RESULTS AND DISCUSSION

A. FIRST 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 4.

TABLE 3. Recognition rates of the system against random forest (RF) (having the number of decision trees between 64–128) instead of using the proposed model (Unit: %). Where WK for walking, RN for running, JM for jumping, SK for skipping, OHW for one hand waving, THW for two hand waving, BD for bending, PJ for place jumping, SM for side movement, CP for clapping, and BX for boxing.

Activities	WK	RN	JM	SK	OHW	THW	BD	PJ	SM	CP	BX
WK	**71**	4	2	3	2	1	5	4	2	3	3
RN	5	**72**	2	4	2	0	5	3	2	4	1
JM	3	5	**68**	2	4	5	1	4	2	4	2
SK	4	6	2	**66**	2	5	4	2	2	5	2
OHW	4	2	3	3	**74**	5	1	1	3	2	2
THW	4	2	5	3	3	**69**	2	4	4	2	2
BD	3	3	4	4	1	5	**68**	4	2	1	5
PJ	3	4	2	3	2	3	1	**76**	2	3	1
SM	2	0	2	4	3	1	4	2	**77**	2	3
CP	3	4	2	5	3	3	4	5	3	**64**	4
BX	3	2	5	4	2	6	5	3	5	3	**62**
Average						**69.7**					

TABLE 4. Recognition rates of the system against support vector machine (SVM) instead of using the proposed model (Unit: %). Where WK for walking, RN for running, JM for jumping, SK for skipping, OHW for one hand waving, THW for two hand waving, BD for bending, PJ for place jumping, SM for side movement, CP for clapping, and BX for boxing.

Activities	WK	RN	JM	SK	OHW	THW	BD	PJ	SM	CP	BX
WK	**70**	5	1	2	4	3	2	4	3	4	2
RN	3	**73**	3	4	3	4	1	3	2	3	1
JM	1	3	**75**	1	3	2	4	2	1	4	3
SK	1	3	2	**78**	2	3	1	4	0	2	4
OHW	4	1	4	2	**77**	0	4	1	3	1	3
THW	3	3	3	3	3	**70**	3	3	3	3	3
BD	3	1	1	3	3	2	**79**	0	4	1	3
PJ	0	3	3	1	1	2	3	**81**	2	4	0
SM	2	2	3	3	3	3	3	3	**72**	3	3
CP	3	4	5	0	3	1	4	1	3	**74**	2
BX	5	2	2	0	6	3	2	4	1	5	**70**
Average						**74.4**					

TABLE 5. Recognition rates of the system against decision tree (DT) instead of using the proposed model (Unit: %). Where WK for walking, RN for running, JM for jumping, SK for skipping, OHW for one hand waving, THW for two hand waving, BD for bending, PJ for place jumping, SM for side movement, CP for clapping, and BX for boxing.

Activities	WK	RN	JM	SK	OHW	THW	BD	PJ	SM	CP	BX
WK	**54**	7	5	4	6	3	5	6	5	3	2
RN	4	**57**	5	4	2	1	5	2	4	0	6
JM	5	6	**60**	3	5	2	3	5	4	2	5
SK	4	5	3	**58**	4	5	6	2	6	3	4
OHW	2	5	4	3	**62**	2	4	6	5	3	4
THW	3	5	2	7	6	**61**	2	5	4	2	3
BD	3	7	5	5	3	2	**57**	6	4	3	5
PJ	6	4	7	3	5	4	3	**55**	5	3	5
SM	3	5	4	6	2	2	5	4	**63**	1	5
CP	6	3	5	4	3	3	2	5	2	**60**	7
BX	4	2	6	5	2	5	3	2	3	6	**62**
Average						**59.0**					

TABLE 6. Recognition rates of the system against artificial neural network (ANN) [42] instead of using the proposed model (Unit: %). Where WK for walking, RN for running, JM for jumping, SK for skipping, OHW for one hand waving, THW for two hand waving, BD for bending, PJ for place jumping, SM for side movement, CP for clapping, and BX for boxing.

Activities	WK	RN	JM	SK	OHW	THW	BD	PJ	SM	CP	BX
WK	**63**	6	3	2	5	3	5	4	3	2	5
RN	4	**72**	3	2	4	1	0	6	3	3	2
JM	2	3	**69**	5	2	4	3	4	2	4	2
SK	4	5	2	**68**	2	3	2	5	3	2	4
OHW	1	5	2	3	**74**	5	2	0	4	1	3
THW	3	4	2	5	1	**71**	2	4	3	4	1
BD	2	3	4	5	1	2	**70**	3	5	3	2
PJ	4	6	3	2	5	1	4	**65**	3	5	2
SM	2	4	1	5	3	4	2	3	**70**	1	5
CP	0	5	6	3	6	2	3	4	2	**66**	3
BX	4	6	5	1	2	5	3	2	4	6	**62**
Average						**68.1**					

It can be seen in Table 1 and Figure 4 that the proposed model showed best performance against naturalistic dataset (collected by depth camera) that has eleven different types of

TABLE 7. Recognition rates of the system against Naive base (NB) instead of using the proposed model (Unit: %). Where WK for walking, RN for running, JM for jumping, SK for skipping, OHW for one hand waving, THW for two hand waving, BD for bending, PJ for place jumping, SM for side movement, CP for clapping, and BX for boxing.

Activities	WK	RN	JM	SK	OHW	THW	BD	PJ	SM	CP	BX
WK	**55**	7	6	6	3	4	2	1	5	6	4
RN	6	**59**	3	4	5	3	2	5	5	2	6
JM	4	3	**62**	5	2	5	4	5	4	2	4
SK	5	2	4	**60**	2	5	4	5	5	3	5
OHW	3	4	2	4	**59**	7	4	5	3	5	4
THW	2	5	4	5	4	**61**	3	5	6	2	3
BD	3	6	5	4	6	3	**57**	5	4	5	2
PJ	5	4	3	5	6	4	7	**54**	5	3	4
SM	3	3	5	6	2	4	3	5	**61**	2	6
CP	2	4	3	5	5	2	3	5	2	**63**	6
BX	1	4	5	3	4	6	2	4	4	6	**61**
Average						**59.3**					

TABLE 8. Recognition rates of the system against logistic regression (LR) instead of using the proposed model (Unit: %). Where WK for walking, RN for running, JM for jumping, SK for skipping, OHW for one hand waving, THW for two hand waving, BD for bending, PJ for place jumping, SM for side movement, CP for clapping, and BX for boxing.

Activities	WK	RN	JM	SK	OHW	THW	BD	PJ	SM	CP	BX
WK	**58**	6	5	4	3	6	2	4	5	3	4
RN	6	**65**	3	4	3	5	2	3	4	2	3
JM	4	6	**67**	2	3	4	5	3	1	2	3
SK	4	6	3	**62**	4	2	3	5	3	5	3
OHW	5	6	5	4	**55**	5	4	6	2	3	5
THW	3	6	4	5	3	**59**	3	6	4	3	4
BD	4	5	6	4	3	4	**64**	3	1	4	2
PJ	5	3	5	4	6	3	5	**57**	3	5	4
SM	2	5	3	6	3	4	2	3	**66**	1	5
CP	2	3	4	5	3	1	5	3	5	**63**	6
BX	4	4	4	4	4	3	3	3	3	3	**65**
Average						**61.9**					

TABLE 9. Recognition rates of the system against hidden Markov model (HMM) instead of using the proposed model (Unit: %). Where WK for walking, RN for running, JM for jumping, SK for skipping, OHW for one hand waving, THW for two hand waving, BD for bending, PJ for place jumping, SM for side movement, CP for clapping, and BX for boxing.

Activities	WK	RN	JM	SK	OHW	THW	BD	PJ	SM	CP	BX
WK	**75**	4	2	3	1	3	4	3	3	2	0
RN	4	**77**	2	3	1	3	2	1	2	3	2
JM	1	3	**80**	0	3	2	3	0	3	1	4
SK	1	1	2	**82**	0	4	2	1	3	2	2
OHW	3	2	3	4	**78**	1	2	1	3	1	2
THW	3	2	4	1	1	**79**	2	3	1	0	4
BD	0	3	2	2	2	2	**81**	2	2	2	2
PJ	2	0	1	3	3	2	1	**82**	0	4	2
SM	3	2	4	2	1	3	1	2	**79**	3	1
CP	1	3	0	1	2	0	2	4	2	**83**	3
BX	0	3	2	3	1	2	0	3	2	4	**80**
Average						**79.6**					

TABLE 10. Recognition rates of the system against conditional random field (CRF) instead of using the proposed model (Unit: %). Where WK for walking, RN for running, JM for jumping, SK for skipping, OHW for one hand waving, THW for two hand waving, BD for bending, PJ for place jumping, SM for side movement, CP for clapping, and BX for boxing.

Activities	WK	RN	JM	SK	OHW	THW	BD	PJ	SM	CP	BX
WK	**86**	2	2	1	2	1	0	2	2	1	1
RN	4	**83**	1	2	1	2	0	2	1	3	1
JM	3	2	**80**	3	2	0	2	2	4	2	2
SK	2	1	0	**82**	1	3	2	1	3	2	3
OHW	2	1	3	2	**78**	4	2	0	3	2	3
THW	1	2	0	2	3	**81**	3	2	1	3	2
BD	1	2	1	0	2	1	**85**	2	1	3	2
PJ	2	2	1	2	3	3	1	**79**	2	2	3
SM	2	2	0	3	2	1	2	3	**78**	4	3
CP	2	0	1	2	3	1	2	3	1	**80**	5
BX	0	2	1	1	2	0	3	0	1	6	**84**
Average						**81.4**					

TABLE 11. Recognition rates of the system against Markov chain (MC) instead of using the proposed model (Unit: %). Where WK for walking, RN for running, JM for jumping, SK for skipping, OHW for one hand waving, THW for two hand waving, BD for bending, PJ for place jumping, SM for side movement, CP for clapping, and BX for boxing.

Activities	WK	RN	JM	SK	OHW	THW	BD	PJ	SM	CP	BX
WK	**74**	4	3	2	5	1	2	3	1	3	2
RN	5	**67**	3	4	2	4	3	5	4	1	3
JM	4	3	**69**	2	4	3	5	3	1	2	4
SK	2	3	4	**71**	2	1	5	2	4	3	3
OHW	3	1	3	4	**77**	0	4	2	1	3	2
THW	0	3	2	2	3	**80**	1	2	3	0	4
BD	3	2	4	1	5	3	**69**	4	3	2	4
PJ	2	1	3	2	4	5	3	**70**	3	4	3
SM	2	5	1	3	2	3	0	4	**76**	3	1
CP	3	1	2	4	2	0	3	2	3	**78**	2
BX	2	2	1	3	2	1	2	0	3	4	**80**
Average						**73.7**					

TABLE 12. Recognition rates of the system against Gaussian mixture model (GMM) instead of using the proposed model (Unit: %). Where WK for walking, RN for running, JM for jumping, SK for skipping, OHW for one hand waving, THW for two hand waving, BD for bending, PJ for place jumping, SM for side movement, CP for clapping, and BX for boxing.

Activities	WK	RN	JM	SK	OHW	THW	BD	PJ	SM	CP	BX
WK	**71**	3	2	4	1	3	2	4	3	4	3
RN	3	**76**	2	4	2	3	1	2	4	1	2
JM	4	3	**68**	4	2	3	4	1	5	3	3
SK	4	5	3	**65**	4	2	3	5	3	1	5
OHW	4	5	3	4	**60**	6	3	4	3	4	4
THW	4	3	1	2	5	**69**	3	4	1	4	4
BD	2	3	1	2	3	5	**72**	3	4	2	3
PJ	4	3	2	3	1	5	3	**70**	3	4	2
SM	2	0	3	4	3	1	3	5	**73**	0	6
CP	4	5	3	4	2	3	4	3	1	**66**	5
BX	4	3	6	2	3	2	4	3	0	6	**67**
Average						**68.8**					

TABLE 13. Recognition rates of the system against hidden conditional random fields (HCRF) instead of using the proposed model (Unit: %). Where WK for walking, RN for running, JM for jumping, SK for skipping, OHW for one hand waving, THW for two hand waving, BD for bending, PJ for place jumping, SM for side movement, CP for clapping, and BX for boxing.

Activities	WK	RN	JM	SK	OHW	THW	BD	PJ	SM	CP	BX
WK	**89**	5	0	2	0	0	0	0	4	0	0
RN	6	**87**	1	2	1	0	0	3	0	0	0
JM	2	4	**90**	0	1	0	1	2	0	0	0
SK	1	2	0	**91**	0	1	3	1	1	0	0
OHW	1	0	3	0	**87**	4	0	2	0	2	1
THW	2	0	1	1	6	**82**	1	5	0	2	0
BD	2	1	0	1	0	1	**92**	1	0	2	0
PJ	1	2	0	2	1	3	0	**88**	0	1	2
SM	5	3	1	4	0	1	0	0	**86**	0	0
CP	0	0	0	0	0	2	1	5	0	**85**	7
BX	0	0	0	0	0	0	2	2	0	6	**90**
Average						**87.9**					

TABLE 14. Recognition rates of the system against recurrent neural network (RNN) [43] instead of using the proposed model (Unit: %). Where WK for walking, RN for running, JM for jumping, SK for skipping, OHW for one hand waving, THW for two hand waving, BD for bending, PJ for place jumping, SM for side movement, CP for clapping, and BX for boxing.

Activities	WK	RN	JM	SK	OHW	THW	BD	PJ	SM	CP	BX
WK	**81**	5	2	1	3	1	3	2	0	1	1
RN	4	**79**	3	1	3	2	1	2	0	4	1
JM	1	2	**76**	3	5	6	2	1	0	3	1
SK	3	0	4	**88**	0	0	3	0	0	2	0
OHW	0	1	0	0	**86**	6	3	1	2	0	1
THW	0	0	0	0	11	**89**	0	0	0	0	0
BD	1	2	2	1	1	2	**80**	3	0	4	4
PJ	0	4	0	0	0	0	3	**89**	4	0	0
SM	0	1	0	0	3	2	1	0	**90**	2	1
CP	0	0	1	0	1	0	0	0	2	**91**	5
BX	0	3	1	2	5	0	3	4	1	4	**77**
Average						**84.2**					

activities. This is because the proposed approach considered the video-sensor observations as the observations of MEMM. We use a modified Viterbi, a machine-learning algorithm, to produce the most probable activity state sequence based on these observations. Moreover, from the most likely state sequence, we predict the activity state is predicted through our stated algorithm.

TABLE 15. Comparison results of the proposed HAR system with the proposed MEMM model against some stat-of-the-art works (Unit: %).

Recent Systems	[44]	[45]	[46]	[47]	[48]	[49]	[50]	[51]	Proposed Approach
Classification Rates	78	83	88	85	91	86	81	72	**96**

B. SECOND EXPERIMENT

This experiment mainly relates to the evaluation of existing approaches/models on the input naturalistic dataset so that we can evaluate our model against them. Our list chosen existing models can be seen in the above description. We present the overall comparison of these systems against MEMM model in Tables 2, 3 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 and 14.

It can be seen from Tables 2, 3 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 and 14 that the proposed recognition model (i.e., MEMM model) achieves higher recognition rate i.e. the recognition ratio of MEMM is higher than the existing models. However, when we exclude the MEMM model, the system's performance is not so commendable. Therefore, given the above evidence, we conclude that our MEMM model has the capability to accurately classify activities in a much more natural environment as shown over the data that is more natural.

C. THIRD EXPERIMENT

Finally, in this experiment, we compare MEMM model with state-of-the-art existing models: [44]–[48]. In this experiment, the collected dataset is employed for which their implementations were borrowed for fair comparison. Moreover, we again use the 10-fold cross-validation as we stated in the earlier section. We present the weighted average recognition rate of existing works against MEMM's weighted average recognition rate in Table 15.

It can be seen in the Table 15 that the proposed human activity recognition system with the proposed maximum entropy Markov model (MEMM) model consistently shows higher recognition rate as opposed to existing models on all of the depth dataset. Hence, the potential of MEMM model is visible and its accuracy is much higher than existing ones and robustly recognizes human activities using video data.

VII. CONCLUSION

Human activity recognition is an important aspect for our daily life communication that can be exploited in many real applications. One major factor that can reduce the accuracy of an HAR system is the high similarity among different activities that can result in low between-class and high within-class variance problems. Recognition module has a great contribution in the performance of a typical HAR system. Most of the previous systems have focused to implement new algorithms for feature extraction and feature selection modules; however, most of them have failed or faced difficulties in recognition stage.

Therefore, in this study, we have proposed a new recognition model based on maximum entropy Markov model (MEMM) to solve the limitations of the existing classifiers. In this model, the states of the human activities are modeled as the states of maximum entropy Markov model (MEMM), in which the video-sensor observations are considered as the observations of MEMM. A modified Viterbi, a machine-learning algorithm, is utilized in order to generate the most probable activity state sequence based on such observations; then, from the most likely state sequence, the activity state is predicted through the proposed algorithm. Unlike most of the existing works, which were evaluated using a heuristic datasets (collected in controlled environment), performance of the proposed system is assessed in a large-scale experimentation using naturalistic dataset in order to show the robustness of the proposed model. The model employed 10−fold cross-validation scheme. The proposed approach outperformed the existing well-known state-of-the-art methods by achieving a weighted average recognition rate of 96.3% across the dataset. The proposed approach is a bit complex time to label the corresponding activity (means complexity wise, the proposed approach is expensive compared to other conventional models); however, our target is to achieve high accuracy instead of considering the complexity issue.

The proposed system has been tested and validated and tested in lab environment. In future, we will implement the proposed model in healthcare domain in order to resolve the privacy concerns. Moreover, in the future work, we will try solve the complexity issue of this work.

REFERENCES

[1] B. M. Demaerschalk, M. L. Miley, T.-E. J. Kiernan, B. J. Bobrow, D. A. Corday, K. E. Wellik, M. I. Aguilar, T. J. Ingall, D. W. Dodick, and K. Brazdys, "Stroke telemedicine," in *Mayo Clinic Proceedings*, vol. 84. Amsterdam, The Netherlands: Elsevier, 2009, pp. 53–64.

[2] Newsletter Article, CISCO. *Telemedicine: Extending Specialist Care to Rural Areas*. Accessed: Oct. 24, 2020. [Online]. Available: https://www.cisco.com/c/dam/en_us/solutions/industries/docs/gov/fedbiz0%81810healthpresence.pdf

[3] I. H. Kraai, M. L. A. Luttik, R. M. de Jong, T. Jaarsma, and H. L. Hillege, "Heart failure patients monitored with telemedicine: Patient satisfaction, a review of the literature," *J. Cardiac Failure*, vol. 17, no. 8, pp. 684–690, Aug. 2011.

[4] B. J. Matuszewski, W. Quan, and L.-K. Shark, "Facial expression recognition," in *Biometrics-Unique and Diverse Applications in Nature, Science, and Technology*. London, U.K.: IntechOpen, 2011.

[5] 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.

[6] 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.

[7] O. Cleve and S. Gustafsson, "Automatic feature extraction for human activity recognitionon the edge," Tech. Rep., 2019.

[8] 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.

[9] A. Jalal, Y.-H. Kim, Y.-J. Kim, S. Kamal, and D. Kim, "Robust human activity recognition from depth video using spatiotemporal multi-fused features," *Pattern Recognit.*, vol. 61, pp. 295–308, Jan. 2017.

[10] S. Kamal, A. Jalal, and D. Kim, "Depth images-based human detection, tracking and activity recognition using spatiotemporal features and modified HMM," *J. Electr. Eng. Technol.*, vol. 11, no. 6, pp. 1857–1862, Nov. 2016.

[11] 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.

[12] D. A. Adama, A. Lotfi, C. Langensiepen, K. Lee, and P. Trindade, "Human activity learning for assistive robotics using a classifier ensemble," *Soft Comput.*, vol. 22, no. 21, pp. 7027–7039, Nov. 2018.

[13] M. H. Siddiqi, M. Alruwaili, A. Ali, S. Alanazi, and F. Zeshan, "Human activity recognition using Gaussian mixture hidden conditional random fields," *Comput. Intell. Neurosci.*, vol. 2019, pp. 1–14, Aug. 2019.

[14] B. Ni, Y. Pei, P. Moulin, and S. Yan, "Multilevel depth and image fusion for human activity detection," *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1383–1394, Oct. 2013.

[15] S. Park, J. Park, M. Al-Masni, M. Al-Antari, M. Z. Uddin, and T.-S. Kim, "A depth camera-based human activity recognition via deep learning recurrent neural network for health and social care services," *Proc. Comput. Sci.*, vol. 100, pp. 78–84, Jan. 2016.

[16] M. H. Siddiqi, "An improved Gaussian mixture hidden conditional random fields model for audio-based emotions classification," *Egyptian Informat. J.*, vol. 22, no. 1, pp. 45–51, Mar. 2021.

[17] A. Ghotkar, P. Vidap, and K. Deo, "Dynamic hand gesture recognition using hidden Markov model by Microsoft Kinect sensor," *Int. J. Comput. Appl.*, vol. 150, no. 5, pp. 5–9, Sep. 2016.

[18] M. Elmezain, A. Al-Hamadi, J. Appenrodt, and B. Michaelis, "A hidden Markov model-based continuous gesture recognition system for hand motion trajectory," in *Proc. 19th Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4.

[19] M. H. Siddiqi, R. Ali, A. M. Khan, E. S. Kim, G. J. Kim, and S. Lee, "Facial expression recognition using active contour-based face detection, facial movement-based feature extraction, and non-linear feature selection," *Multimedia Syst.*, vol. 21, no. 6, pp. 541–555, Nov. 2015.

[20] M. H. Siddiqi, M. Alruwaili, and A. Ali, "A novel feature selection method for video-based human activity recognition systems," *IEEE Access*, vol. 7, pp. 119593–119602, 2019.

[21] 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.

[22] J. Basavaiah and C. M. Patil, "Robust feature extraction and classification based automated human action recognition system for multiple datasets," *Int. J. Intell. Eng. Syst.*, vol. 13, no. 1, pp. 13–24, 2020.

[23] S.-R. Ke, H. L. U. Thuc, Y.-J. Lee, J.-N. Hwang, J.-H. Yoo, and K.-H. Choi, "A review on video-based human activity recognition," *Computers*, vol. 2, pp. 88–131, Jun. 2013.

[24] H. H. Ali, H. M. Moftah, and A. A. A. Youssif, "Depth-based human activity recognition: A comparative perspective study on feature extraction," *Future Comput. Informat. J.*, vol. 3, no. 1, pp. 51–67, Jun. 2018.

[25] M. H. Siddiqi, R. Ali, A. M. Khan, Y. T. Park, and S. Lee, "Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields," *IEEE Trans. Image Process.*, vol. 24, no. 4, pp. 1386–1398, Apr. 2015.

[26] S. H. Kung, M. A. Zohdy, and D. Bouchaffra, "3D HMM-based facial expression recognition using histogram of oriented optical flow," *Trans. Mach. Learn. Artif. Intell.*, vol. 3, no. 6, p. 42, Dec. 2015.

[27] M. A. Khan, K. Javed, S. A. Khan, T. Saba, U. Habib, J. A. Khan, and A. A. Abbasi, "Human action recognition using fusion of multiview and deep features: An application to video surveillance," *Multimedia Tools Appl.*, pp. 1–27, Mar. 2020.

[28] S. Mohammed, "What are the disadvantages of Naive Bayes," Tech. Rep., May 2020.

[29] K. K. Htike, O. O. Khalifa, H. A. Mohd Ramli, and M. A. M. Abushariah, "Human activity recognition for video surveillance using sequences of postures," in *Proc. 3rd Int. Conf. e-Technol. Netw. Develop. (ICeND)*, Apr. 2014, pp. 79–82.

[30] W. Jia, D. Zhao, T. Shen, Y. Tang, and Y. Zhao, "Study on optimized Elman neural network classification algorithm based on PLS and CA," *Comput. Intell. Neurosci.*, vol. 2014, pp. 1–13, Aug. 2014.

[31] S. Sehgal, "Human activity recognition using BPNN classifier on HOG features," in *Proc. Int. Conf. Intell. Circuits Syst. (ICICS)*, Apr. 2018, pp. 286–289.

[32] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, no. 1, Jun. 2005, pp. 886–893.

[33] N. M. Nawi, A. Khan, and M. Z. Rehman, "A new back-propagation neural network optimized with cuckoo search algorithm," in *Proc. Int. Conf. Comput. Sci. Appl.*, Springer, 2013, pp. 413–426.

[34] N. A. Hamid, N. M. Nawi, R. Ghazali, and M. N. M. Salleh, "Solving local minima problem in back propagation algorithm using adaptive gain, adaptive momentum and adaptive learning rate on classification problems," *Int. J. Modern Phys., Conf. Ser.*, vol. 9, pp. 448–455, Jan. 2012.

[35] W. Leigh, R. Hightower, and N. Modani, "Forecasting the New York stock exchange composite index with past price and interest rate on condition of volume spike," *Expert Syst. Appl.*, vol. 28, no. 1, pp. 1–8, Jan. 2005.

[36] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification," in *Proc. OTM Confederated Int. Conf. Move Meaningful Internet Syst.*, Springer, 2003, pp. 986–996.

[37] J. Ouyang, H. Luo, Z. Wang, J. Tian, C. Liu, and K. Sheng, "FPGA implementation of GZIP compression and decompression for IDC services," in *Proc. Int. Conf. Field-Program. Technol.*, Dec. 2010, pp. 265–268.

[38] 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.

[39] M. H. Siddiqi, "Accurate and robust facial expression recognition system using real-time YouTube-based datasets," *Int. J. Speech Technol.*, vol. 48, no. 9, pp. 2912–2929, Sep. 2018.

[40] D. Jurasky and J. H. Martin, "Speech and language processing: An introduction to natural language processing," in *Computational Linguistics and Speech Recognition*. Upper Saddle River, NJ, USA: Prentice-Hall, 2000.

[41] M. H. Siddiqi, M. G. R. Alam, C. S. Hong, A. M. Khan, and H. Choo, "A novel maximum entropy Markov model for human facial expression recognition," *PLoS ONE*, vol. 11, no. 9, Sep. 2016, Art. no. e0162702.

[42] N. Sairamya, L. Susmitha, S. T. George, and M. Subathra, N. Sairamya, L. Susmitha, S. T. George, and M. Subathra, "Hybrid approach for classification of electroencephalographic signals using time–frequency images with wavelets and texture features," in *Intelligent Data Analysis for Biomedical Applications*. Amsterdam, The Netherlands: Elsevier, 2019, pp. 253–273.

[43] D. Singh, E. Merdivan, I. Psychoula, J. Kropf, S. Hanke, M. Geist, and A. Holzinger, "Human activity recognition using recurrent neural networks," in *Proc. Int. Cross-Domain Conf. Mach. Learn. Knowl. Extraction*. Springer, 2017, pp. 267–274.

[44] A. Bagate and M. Shah, "Human activity recognition using RGB-D sensors," in *Proc. Int. Conf. Intell. Comput. Control Syst. (ICCS)*, May 2019, pp. 902–905.

[45] H. Yao, M. Yang, T. Chen, Y. Wei, and Y. Zhang, "Depth-based human activity recognition via multi-level fused features and fast broad learning system," *Int. J. Distrib. Sensor Netw.*, vol. 16, no. 2, 2020, Art. no. 1550147720907830.

[46] S. Nehra and J. L. Raheja, "Unobtrusive and non-invasive human activity recognition using Kinect sensor," in *Proc. Indo Taiwan 2nd Int. Conf. Comput., Anal. Netw. (Indo-Taiwan ICAN)*, Feb. 2020, pp. 58–63.

[47] 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.

[48] 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. 2017.

[49] A. Nadeem, A. Jalal, and K. Kim, "Accurate physical activity recognition using multidimensional features and Markov model for smart health fitness," *Symmetry*, vol. 12, no. 11, p. 1766, Oct. 2020.

[50] J. Sung, C. Ponce, B. Selman, and A. Saxena, "Unstructured human activity detection from RGBD images," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2012, pp. 842–849.

[51] J. Sung, C. Ponce, B. Selman, and A. Saxena, "Human activity detection from RGBD images," in *Proc. Workshops 25th AAAI Conf. Artif. Intell.*, 2011, pp. 1–8.

IBRAHIM ALRASHDI received the B.S. degree in computer science and information from Jouf University, Saudi Arabia, in 2009, the M.S. degree in computer science from Western Illinois University, IL, USA, in 2013, and the Ph.D. degree in computer science and informatics from Oakland University, MI, USA, in 2019. He is currently an Assistant Professor with the Department of Computer Science, College of Computer and Information Sciences, Jouf University. His research interests include the Internet of Things, cybersecurity, and artificial intelligence.

MUHAMMAD HAMEED SIDDIQI received the Bachelor of Computer Science degree (Hons.) from Islamia College (Chartered University) Peshawar, Khyber Pakhtunkhwa, 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. He was an Assistant Professor with the Department of Computer Science, Jouf University, Sakaka, Saudi Arabia, from September 2016 to October 2020, where he has been working as an Associate Professor, since November 2020. He was also a Postdoctoral Research Scientist with the Department of Computer Science and Engineering, Sungkyunkwan University, Suwon, from March 2016 to August 2016. He was a Graduate Assistant with Universiti Teknologi Petronas, Malaysia, from 2008 to 2009. 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.

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 in 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.

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.

MOHAMMAD AZAD received the Ph.D. degree in computer science from the King Abdullah University of Science and Technology, Saudi Arabia. He is currently working as an Assistant Professor with the Department of Computer Science, Jouf University, Saudi Arabia. He is the author or coauthor of one research book published by Springer and over 38 international journal and conference papers. He is also acting as a reviewer of many international journals.