COMPLEX HUMAN ACTIVITY RECOGNITION USING GAUSSIAN POLAR COORDINATE AND SEQUENCE LIKELIHOOD SUPPORT CLASSIFIER

BagavathiLakshmi[a], S.Parthasarathy[b,*]

[a]Research and Development Centre, Bharathiar University, Coimbatore, India

[b]Department of Computer Applications, Thiagarajar College of Engineering, Madurai India

Abstract

In last few years, complex human activity recognition acquired sizeable research attentions. This is due to extensive scope of pattern recognition owing to its noteworthy applications such as robot learning, surveillance, user interface design and so on. For example, activity recognition systems can be applied in human robot interaction in industry to improve their assembly tasks. During the past few years, orientation variations and joint location sequence is widely used for capturing human sequence after extracting discriminative features from frames. However, robust features with complex recognition of human activity were not focused. To address this issue, a Gaussian Polar Coordinate with Sequence Likelihood Support Classifier (GPC-SLSC) framework for complex activity recognition is investigated. To start with, in this work, differentiation between the background and foreground frame are made from raw data using Gaussian Foreground Frame Detection. The obtained features are further processed using Polar Coordinate Feature Extraction model to make them more robust. Finally, the features are trained with Sequence Likelihood Support Vector Machine Classifier for successful complex human activity recognition. The proposed framework is compared with the traditional human activity recognition approaches such as Compressive sensing dictionary-based approach and Long Short Term Memory (LSTM) Recurrent Neural Network where it outperformed them. The results of experiments show that the framework recognizes complex activities involving maximum accuracy with minimum time and complexity.

Keywords: Gaussian, Frame Detection, Polar Coordinate, Feature Extraction, Sequence Likelihood, Support Classifier

1. Introduction

With incredible contributions in worldwide computing, Human Activity Recognition (HAR) has become a distinguished research area. Research persons use these HAR models as a means to obtain information about peoples' behavior. Also, in the past few years, human activity recognition acquired sizeable research observations from an extensive scope of pattern recognition due to its distinguished applications such as smart home health care.

Compressive sensing dictionary-based approach [1] involved the design of unobtrusive and robust system to recognize human activities using machine learning algorithms. First unsupervised subspace decomposition was applied to the input data to extract the features avoiding noisy and streaming data. Next, to achieve robust activity recognition, the properties of sparse coefficients were used.

Finally, more promising representation of activities was evolved and also preserved rich information with higher accuracy level. Despite improvement found in mean recognition rate and accuracy, more robust features were not focused and also complex activity recognition was not performed. To address this issue by identifying the robust features, a Polar Coordinate Robust Feature Extraction model is used. Robust features are said to be extracted here using polar coordinate values.

A deep learning model that automatically learned to classify human activities without using any prior knowledge, called, Long Short Term Memory (LSTM) Recurrent Neural Network was investigated in [2]. The objective of LSTM Recurrent Neural Network was to identify both simple and complex activities with the help of sensor data. Here, Long Short-Term Memory classifier was adopted to classify human activities such as cooking, bathing and sleeping. Further Recurrent Neural Network was applied to the classified data for human activity recognition.

The results proved that the deep learning based approach used provided significant improvement in performance and ensured increased accuracy. Despite improvement found in performance and accuracy, complex activity recognition for reducing variance on predictions remained unaddressed. To address this issue, a complex human activity recognition framework to minimize variance to control errors is investigated using Maximum Sequence Likelihood Support Vector Machine Classifier.

This paper proposes two alternatives to the complex activity recognition and hence, the conventional human activity recognition to improve the detection rate with minimum time and complexity. To address the limitation of conventional human activity recognition, initially, foreground pose detection is performed. Next, with the detection foreground pose, robust features are extracted. Finally, by applying machine learning algorithm, complex human activity recognition is performed.

The main contribution of this paper is the introduction of Gaussian Foreground Frame Detection with a two dimensional and average two dimensional matrix representation which differentiates between the foreground and background frame. Also by investigating Maximum Sequence Likelihood Support Vector Machine Classifier, robust features with complex human activity recognition are made through maximum likelihood of features. This choice is motivated by the benefit of compressive sensing and dictionary-based approach in device-free activity recognition [1] and increasing variations on predictions by long short term memory recurrent neural network [2]. Quantitative measure to evaluate the robustness and complex activity recognition is also made and performance comparison is measured on a common recognition framework on two different benchmark human action recognition datasets.

The rest of this paper is organized as follows. Section 2 reviews the use of certain human activity recognition in the existing methods. Section 3 explains the concept of Gaussian Polar Coordinate with Sequence Likelihood Support Classifier (GPC-SLSC) framework for complex activity recognition. Section 4 describes the experimental setting. Section 5 discusses with different parameters. Finally, Section 7 concludes the paper.

2. Related works

A robust human activity recognition system using Deep Belief Network-based approach (DBN-based approach) was investigated in [4]. In DBN-based approach, smart phone inertial sensors were used to obtain the input data. With the obtained input, efficient features were initially extracted that included mean, median, autoregressive coefficients, etc. Besides to make them robust, the features were then processed using a Kernel Principal Component Analysis (KPCA) and Linear Discriminant Analysis (LDA). Finally, the obtained robust features were trained using Deep Belief Network (DBN) for efficient human activity recognition, therefore improving the mean recognition rate and overall accuracy.

Multiple dictionaries of action primitives were analyzed in [5] at different resolutions. However, measures to determine the best feature combination was not investigated. To address this issue, feed forward artificial neural network with k nearest neighbor and decision tree was designed in [6] with the objective of improving human activity recognition.

In the recent few years, activity recognition using wireless signal has received significant attention. Device-free movement recognition was investigated in [7] leveraging radio frequency to obtain high accuracy of human activity recognition. On contrary, to analyze hand gestures, normalization of features under root mean square curve were performed in [8], therefore improving substantial amount of human-computer interaction.

To recognition activity in real time environment, deep learning approach was presented in [9] addition to spectral domain pre-processing. This not only resulted in the improvement of accuracy but also reduced the computation time involved in recognition. Yet another matrix model using low rank two dimensional and 3-D Hankel structures were introduced in [10] to perform recognition of human activity with robust classification accuracy.

With the increase in the development of electronics, certain wearable devices, namely, Fitbit, Nike FuelBand, are swiftly becoming the most prominent part of our daily lives. One of the key services provided by these wearable devices is continuous movement tracking of users with the help of inbuilt precision accelerometers.

A Human Activity Recognition from Kinetic Energy Harvesting (HARKE) was investigated in [11] with the resultant accuracy found to be improved and quantify power saving. Yet another, context associative hierarchy memory was designed in [12] to integrate human level concepts to extensible computational framework using concept hierarchy. Contour points and multi-view key poses were employed in [13] to make various pose representation and learning actions for human action recognition, resulting in high and stable success rate.

A combination of Spatio Temporal Interest Points (STIP) and Bag Of Words (BOW) model were integrated in [14] for efficient human motion recognition in real video data. Human activity recognition is not considered to be an easy task. This is due to the fact that actions do not involve any formal definition.

Actions vary highly for several human with different styles and different human body sizes between persons to persons. An Extreme Learning Machine (ELM) algorithm was investigated in [15] with the objective of dynamic recognition of human activity. Yet another integration of gabor filtering and bag of words model was introduced in [16] based on the Laplacian pyramid to demonstrate the effectiveness of human activity recognition.

Hull convexity defect features were utilized in [17] for recognition human activity by applying Principal Component Analysis and encoded neural networks. This not only improved the accuracy rate but also minimized the computational complexity involved in human recognition.

Three different machine learning techniques namely, K Means Clustering, Support Vector Machine and Hidden Markov Models were employed in [18] to recognize the activities in real time environment. A review on human activity recognition methods were investigated in [19]. Interval temporal syntactic model was used in [20] to recognize complex human activity.

In the analyzed papers, the authors employ different mechanisms for human activity recognition there are no changes on-the-fly in the robust feature extraction and complex human activity recognition. In this way, noise and recognition rate is said to be compromised and hence has to be explored. The work proposed in this paper takes account of robust feature extraction and complex recognition of human activities using Gaussian Polar Coordinate with Sequence Likelihood Support Classifier (GPC-SLSC) framework, which is discussed in the forthcoming sections.

3. **Methodology**

In this section, a detailed description of our Gaussian Polar Coordinate with Sequence Likelihood Support Classifier (GPC-SLSC) framework for complex activity recognition is provided. Figure 1 illustrates the process of training a classifier to recognize complex human activities. The GPC-SLSC framework consists of three different parts. They are frame detection, robust feature extraction and complex human activity recognition.

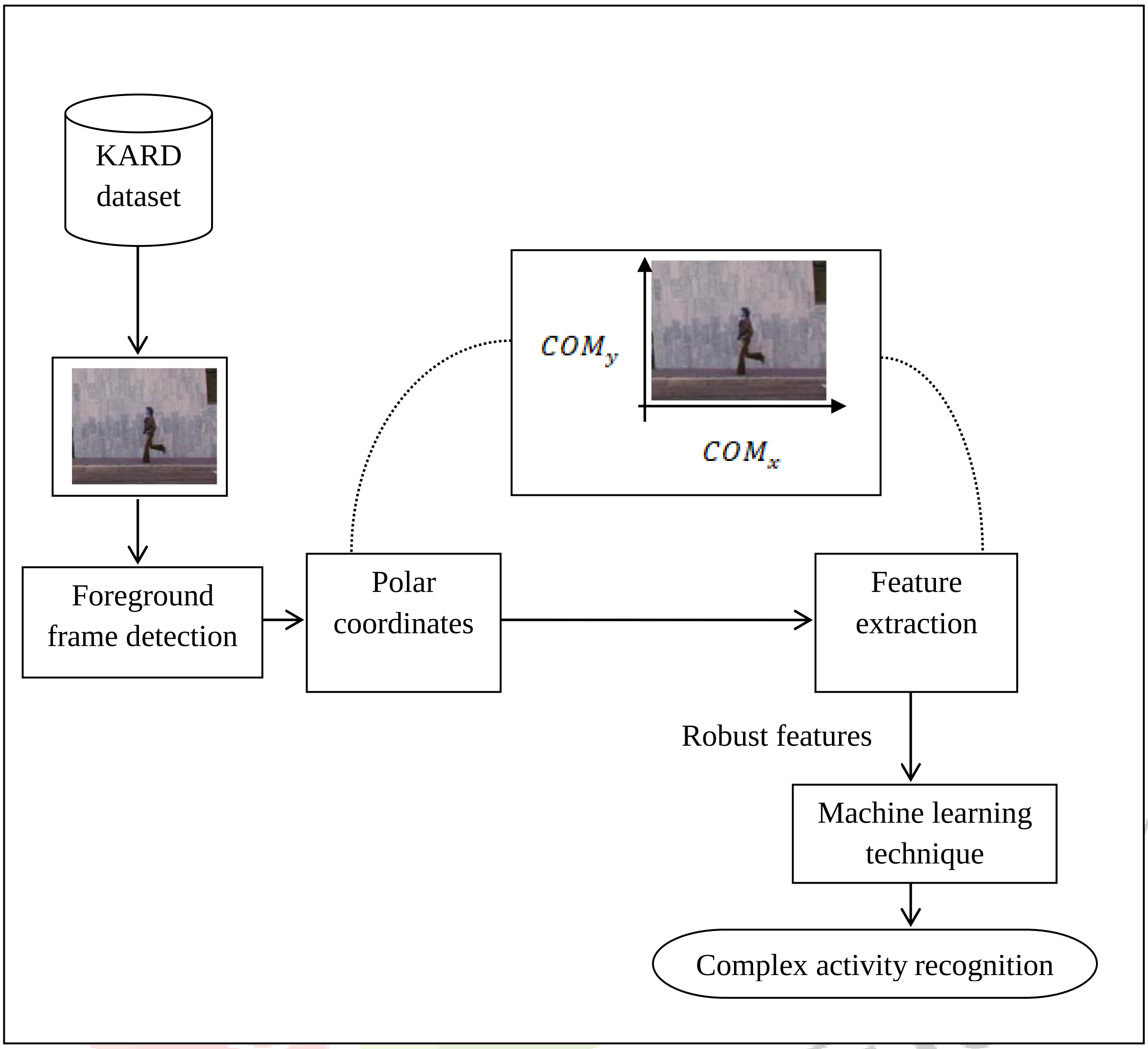
KARD dataset
COM_y
COM_x
Foreground frame detection
Polar coordinates
Feature extraction
Robust features
Machine learning technique
Complex activity recognition

Figure 1 Block diagram of proposed framework

As illustrated in the figure, first, with Kinect Activity Recognition Dataset (KARD) [3] given as input, Gaussian Foreground Frame Detection is applied to detect the foreground frame or the actual region of interest. Next, Polar Coordinate Feature Extraction is applied to extract the robust features. Finally, a Sequence Likelihood Support Vector Machine Classifier algorithm is applied to the robust features extracted to recognize complex human activity.

3.1 Gaussian Foreground Frame Detection

Let us consider Kinect Activity Recognition Dataset (KARD) [3] dataset provided as input to the proposed framework. Here, the KARD dataset consists of several videos ‘$V = v_1, v_2, \dots, v_n$’, split into frames ‘$F = f_1, f_2, \dots, f_n$’. To identify only human poses and reducing noise, instead of comparing two frames, the proposed framework compares each foreground frame with the background frame. Here, the foreground frame refers to the human pose and other objects present in the frame are considered as the background frame. As the objective lies only in recognizing complex human activity, the background frame is discarded. This is performed using Gaussian Foreground Frame Detection. Here, Gaussian refers to the frames. The Gaussian (i.e. frames) assumes all the data points or frames.

An empty two dimensional matrix ‘$2DM$’ for a Gaussian (i.e. frame) ‘f’ is created as follows. To start with, all its values are assigned with zeros. The mathematical representation of ‘$2DM$’ with rows ‘r’ and columns ‘c’ is given as below.

$$2DM = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1r} \\ d_{21} & d_{22} & \dots & d_{2r} \\ \dots & \dots & \dots & \dots \\ d_{c1} & d_{c2} & \dots & d_{cc} \end{bmatrix} \qquad (1)$$

Then, the averages of above two dimensional matrix ‘$2DM$’ is mathematically evaluated as given below.

$$A2DM = \begin{bmatrix} d_{11}/n & d_{12}/n & \dots & d_{1r}/n \\ d_{21}/n & d_{22}/n & \dots & d_{2r}/n \\ \dots & \dots & \dots & \dots \\ d_{c1}/n & d_{c2}/n & \dots & d_{cc}/n \end{bmatrix} \qquad (2)$$

Next, to differentiate between background and foreground Gaussians (i.e. frames), the averages of overall Gaussians (i.e. frames) are mathematically evaluated as given below.

$$DM = \sum_{i=1}^{r} \sum_{j=1}^{c} d_{ij} \qquad (3)$$

$$ADM = \sum_{i=1}^{r} \sum_{j=1}^{c} d_{ij}\ /n \qquad (4)$$

From the above equation () and (), the differentiation between the background and foreground frame are made. The pseudo code representation of Gaussian Frame Detection is given below.

```
Input: Input: videos ‘$V = v_1, v_2, \dots, v_n$’, frames ‘$F = f_1, f_2, \dots, f_n$’
Output: Foreground pose frames ‘$FF = ff_1, ff_2, \dots, ff_n$’
Begin
    For each video ‘V’ split into frames ‘F’
        Repeat
            Obtain two dimensional matrix representation using equation (1)
            Obtain averages of above two dimensional matrix using equation (2)
            Measure averages of overall frames using equation (3) and (4)
            If ‘$ADM$’ > ‘$DM$’
                Obtained frames are assigned as foreground ‘$FF$’
            Else
                Obtained frames are assigned as background
                Discard it
            End if
        Until (all ‘$n$’ frames are executed)
    End for
End
```

Algorithm 1 Gaussian Frame Detection algorithm

As given in the above Gaussian Frame Detection algorithm, for each video frame obtained from KARD dataset, the pixel values of frames are represented in a two dimensional matrix. Followed by the representation, the average of it is measured. Finally, the background and foreground frame are differentiated through Gaussian comparison to extract the actual features (i.e. human pose). In this way, the unwanted background frames are discarded, therefore reducing the noise.

3.2 Polar Coordinate Feature Extraction model

With the obtained Foreground pose frames ‘$FF = ff_1, ff_2, \dots, ff_n$’ as input, the second part represents the actual feature extraction. For the design of robust feature extraction, Foreground pose frames ‘$FF = ff_1, ff_2, \dots, ff_n$’ obtained from Gaussian Frame Detection algorithm is considered as input. With the objective of extracting robust feature, a polar coordinate model is designed. Figure 2 given below shows the flow diagram of Polar Coordinate Robust Feature Extraction model.

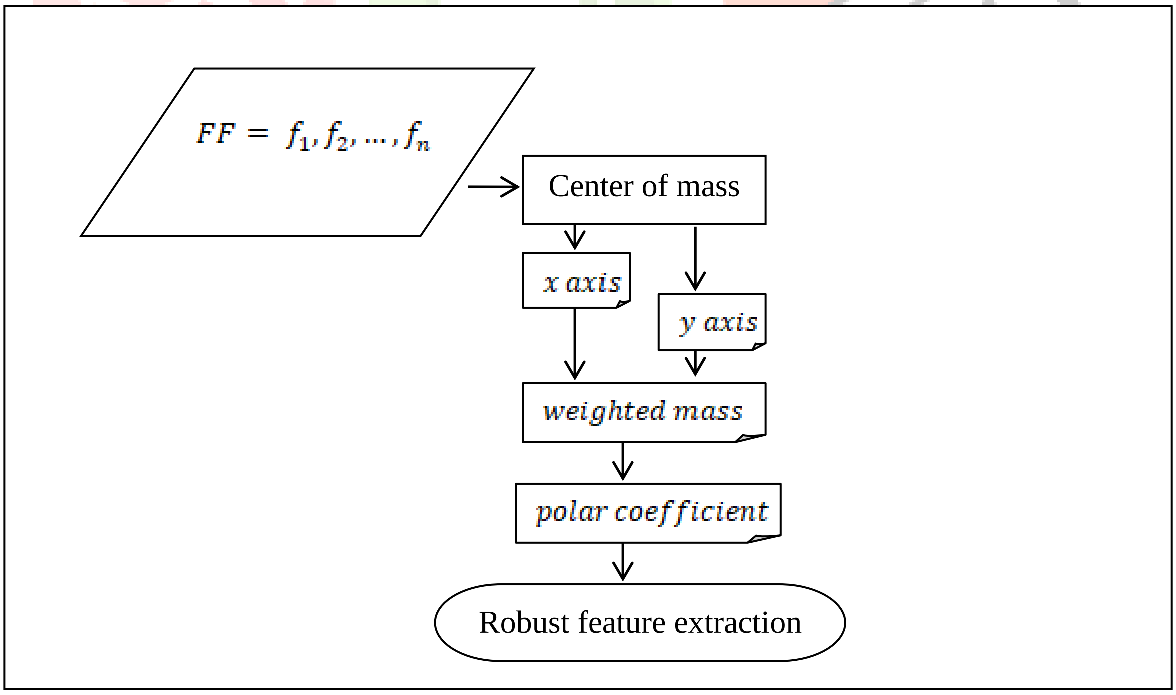

Figure 2 Flow diagram of Polar Coordinate Robust Feature Extraction model

As shown in the figure, let us consider a prototype ‘$FF = \{\alpha_r \mid i = 1, 2, \ldots, j_i\}$’, where ‘$\alpha_r$’ corresponds to referral points and ‘j_i’ corresponds to joints for each foreground pose frame ‘FF’, obtained from videos or a small sequence of poses. Then, the objective remains in extracting the robust features from small sequence of poses. The sequence of poses is obtained through foreground pose frame that corresponds to referral points and joints for complex human activity recognition. Here, ‘$\alpha_r = (p_r, q_r)$’ is selected as the center of mass.

The center of mass ‘COM’ here refers to the position defined relative to a foreground pose frame or frames. Besides, it represents the average position of all parts and weighted based on their masses. It is mathematically formulated separately for object (i.e. sequence of poses) positioning along the x axis and objects (i.e. sequence of poses) positioning along the y axis and is mathematically formulated as given below.

$$COM_x = \frac{m_1 x_1 + m_2 x_2 + \ldots + m_n x_n}{m_1 + m_2 + \ldots + m_n} \qquad (5)$$

$$COM_y = \frac{m_1 y_1 + m_2 y_2 + \ldots + m_n y_n}{m_1 + m_2 + \ldots + m_n} \qquad (6)$$

From the above equation (5) and (6), ‘COM_x’ and ‘COM_y’ represents the center of mass along the ‘x’ axis and ‘y’ axis respectively with ‘m_1’, ‘m_2’, the corresponding masses. Furthermore, to extract efficient and robust features, the ‘ith’ joint is represented as ordered set ‘OS’ of features and is mathematically formulated as given below.

$$OS = (v_i, PC_i, \beta_i, w_i) \qquad (7)$$

From the above equation (7), the ordered set ‘OS’ in the proposed framework for extracting efficient and robust features comprises of the joint video identifier ‘v_i’, with ‘PC_i’ and ‘β_i’ representing the polar coordinates of position of ‘ith’ joint vector. Here, the polar coordinate system for extracting robust features comprises of a two dimensional coordinate system, where the referral point is referred to as the pole. The mathematical formulation of the polar coordinates is as given below.

$$PC_i = \sqrt{(p_i - p_r)^2 + (q_i - q_r)^2} \qquad (8)$$

$$\beta_i = tan^{-1} \left[\frac{q_i - q_r}{p_i - p_r}\right] \qquad (9)$$

From above equation (8) and (9), ‘p_r’ and ‘q_r’ represents the referral points of ‘p_i’ and ‘$q_i th$’ joint vector. Finally, ‘w_i’ in (7) represents the weight allocated to the joint, where the center of mass for the corresponding entire foreground pose frame is split into two portions, according to the upper and lower parts of the body. This is turn permits to improve the acuteness of the features being extracted for recognizing robust human activity that involve only specific parts of the body.

For example, with the actions involving running, only the lower parts of the body is of high interest and therefore the upper parts of the body are not included for further processing. Finally, the value of foreground pose frame to be extracted is obtained by combining the scores of joints using weighted geometric mean and is mathematically formulated as given below.

$$RF = WGM = \prod_{i=1}^{n} V_i^{w_i \frac{1}{\sum_{i=1}^{n} w_i}} \quad (10)$$

From the above equation (10), the weighted geometric mean 'WGM' is evaluated using the single video vector 'V_i' in the test sample, weight of the single video vector 'w_i' in the test sample and the sum of the weights 'w_i' respectively. The pseudo code representation of Weighted Geometric Feature Extraction is given below.

```
Input: Foreground pose frames '$FF = ff_1, ff_2, \dots, ff_n$', videos '$V = v_1, v_2, \dots, v_n$', frames '$F = f_1, f_2, \dots, f_n$'
Output: robust features extracted '$RF = rf_1, rf_2, \dots, rf_n$'
Begin
    For each Foreground pose frames '$FF$'
        Measure center of mass along x axis and y axis using equation (5) and (6)
        Form the ordered set of features using equation (7)
        Measure the polar coordinates using equation (8) and (9)
        Measure the weighted geometric mean using equation (10)
    End for
End
```

Algorithm 1 Weighted Geometric Feature Extraction algorithm

As given in the above algorithm, for each Foreground pose frames 'FF', center of mass along the x axis and y axis are measured separately so that the actual actions can be recognized. Next, the polar coordinates are measured from referral point to the angle from referral point. Finally, the scores for each video are obtained by measuring the weighted geometric mean, so that robust features are extracted.

3.3 Maximum Sequence Likelihood Support Vector Machine Classifier

With the extracted robust features, finally, recognition of complex human is said to be performed using Maximum Sequence Likelihood Support Vector Machine Classifier. Here, complex human activity recognition refers to the inferring of concurrent activities. Figure shows the activity diagram of Maximum Sequence Likelihood Support Vector Machine Classifier.

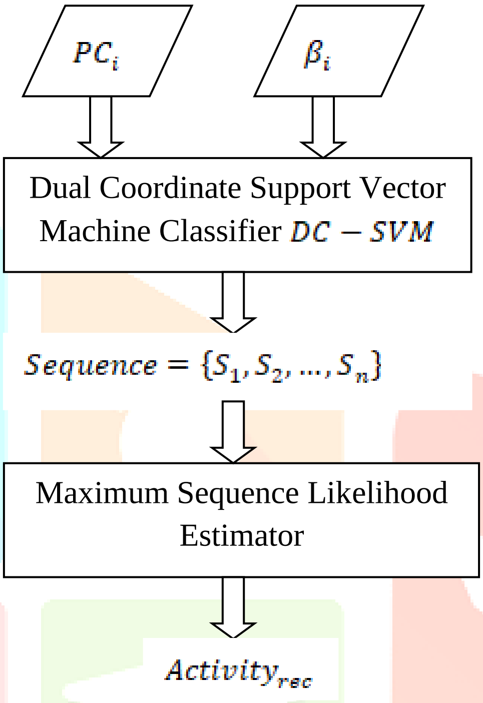
PC_i
β_i
Dual Coordinate Support Vector Machine Classifier $DC - SVM$
$Sequence = \{S_1, S_2, \dots, S_n\}$
Maximum Sequence Likelihood Estimator
$Activity_{rec}$

Figure 3 Activity diagram of Maximum Sequence Likelihood Support Vector Machine Classifier

As illustrated in the figure, we computed the polar coordinates 'PC_i', 'β_i' and therefore used them as two features. Then, the correlation among the two features is mathematically formulated as given below.

$$PC_i = PC_1\ [rf_1], PC_2\ [rf_2], \dots, PC_n\ [rf_n] \tag{11}$$

$$\beta_i = \beta_1[rf_1], \beta_2[rf_2], \dots, \beta_n[rf_n] \tag{12}$$

From above equation (11) and (12), the correlation among the two features 'PC_i' and 'β_i', with respect to the robust feature extracted 'rf_i' are obtained. With this as the input, complex human activity are analyzed. To this, the sequences are given as input and are mathematically formulated as given below.

$$Sequence = \{S_1, S_2, \dots, S_n\} \tag{13}$$

$$= DC - SVM\ (PC_i, \beta_i\ ) \cup Prob\ (S_j \mid Cl_k), j = 1,2,..,n \qquad (14)$$

$$= Bayes\ \{Sequence\ |Cl_k\ \} \qquad (15)$$

From the above equation (13), (14) and (15), 'n' sequences 'S_n' are provided for recognizing complex human activity. As dual coordinates i.e. 'PC_i' and 'β_i' are included for design both are fed and combined with the 'j' sequences 'S_j', for forming 'k' classes 'Cl_k'. Next, a Bayes classifier is applied to obtain all the likelihoods. Finally, maximum likelihood is applied to recognize the complex human activity and is mathematically formulated as given below.

$$Activity_{rec} = MAX\ \{Prob\ (Sequence = S_j\ |Cl_k\ )\} \qquad (16)$$

From the above equation (16), complex human activity recognition '$Activity_{rec}$' is performed based on the maximum likelihood of probability sequences '$MAX\{Prob\ (Sequence)\}$'. The pseudo code representation of Sequence Likelihood Support Vector Machine Classifier is given below.

```
Input: robust features '$RF = rf_1, rf_2, \dots, rf_n$', Sequence $\{S_1, S_2, \dots, S_n\}$,
Output: Complex human activity recognition
Begin
    For each robust features '$RF$'
        Obtain correlation among two features using equation (13) and (14)
        Measure maximum sequence likelihood using equation (16)
    End for
End
```

Algorithm 3 Sequence Likelihood Support Vector Machine Classifier

As given in the above algorithm, the polar coordinates were normalized with respect to the robust features obtained. Then, the dual coordinate estimator was used to build the classification model. Followed by which, the maximum likelihood estimator was evolved for each sequences. Finally, for these sequences, the synthesized data were classified using the dual coordinate estimator and the recognized result was obtained.

4. Performance evaluation

Based on the design explained above, our complex human activity recognition system is implemented in JAVA platform on Windows 7. Several experiments were conducted to evaluate the performance of our proposed complex human activity recognition system. In the experiments, two different datasets are used: Weizmann human motion dataset [21] and UIUC action dataset [22].

The WEIZMANN dataset as provided by Blank comprises of 93 video sequences. The 93 video sequences show nine different people. Nine different people perform 10 different actions. The actions included are, running, walking, skipping, jump-jack, jump-forward-on-two-legs, jump-in-place-ontwo- legs, gallop-sideways, wave-two-hands, wave-one-hand, and bending.

On the other hand, the UIUC action dataset as created by the University of Illinois at Urbana-Champaign (UIUC) in 2008 is specifically designed for human activity recognition. The activities included in the UIUC action dataset are walking, running, jumping, waving, jumping jacks, clapping, jumping from sit up, raising one hand, stretching out, turning, sitting to standing, crawling, pushing up, and standing to sitting.

For two datasets, 10 video sequences with 10 different actions and so total of 100 video sequences are taken to conduct experiments. The experiments are repeated ten times. The performance of the proposed framework, Gaussian Polar Coordinate with Sequence Likelihood Support Classifier (GPC-SLSC) framework for complex activity recognition is compared with the two existing methods, Compressive sensing dictionary-based approach [1] and Long Short Term Memory (LSTM) Recurrent Neural Network [2]. In order to evaluate the performance of complex activity recognition, four different parameters are tested with, computational time, precision, recall and human activity recognition accuracy.

Complex human activity recognition time refers to the time consumed while recognizing complex human activity. It is mathematically formulated as given below.

$$CHAR_t = \sum_{i=1}^{n} S_i * Time\ [Activity_{rec}] \qquad (17)$$

From the above equation (17), the complex human activity recognition time '$CHAR_t$' is measured on the basis of the number of sequences provided as input and the time taken for activity recognition based on maximum likelihood function. Here, the maximum likelihood function refers to the maximum likelihood of probability sequences with respect to different classes, to recognize the complex human activity.

Complex human activity recognition accuracy measures the ratio of successful rate of recognition to the sequences provided as input. It is mathematically formulated as given below.

$$CHAR_A = \sum_{i=1}^{n} \frac{SSR}{S_i} * 100 \qquad (18)$$

From the above equation (18), the complex human activity recognition accuracy '$CHAR_A$' is evaluated using successful sequence recognition 'SSR' and sequences 'S_i' considered for experimentation. Precision refers to the ratio of the number of relevant sequences retrieved to the total number of irrelevant and relevant sequences retrieved and are expressed as a percentage.

$$P = \frac{A}{A+C} * 100 \qquad (19)$$

From the above equation (19), precision 'P' is measured with respect to the number of relevant sequences retrieved 'A', to the summation of relevant sequences retrieved and number of irrelevant sequences retrieved '$A + C$'. On the other hand, recall measures the ratio of the number of relevant sequences retrieved to the number of relevant records in the dataset and are expressed as percentage (%).

$$R = \frac{A}{A+B} * 100 \qquad (20)$$

From the above equation (20), recall 'R' is measured with respect to number of relevant sequences retrieved 'A' to the summation of relevant sequences retrieved and number of relevant sequences not retrieved '$A + B$'.

5. Discussion

To analyze the performance of our complex human activity recognition system, four different experiments were conducted using Weizmann human motion and UIUC action datasets. Four different experiments were complex human activity recognition time, complex human activity recognition accuracy, precision and recall. Comparison was made with two different recognition methods, Compressive sensing dictionary-based approach [1] and Long Short Term Memory (LSTM) Recurrent Neural Network [2].

5.1 Scenario 1: Impact of complex human activity recognition time

In the first experiment, complex human activity recognition time is performed and comparison made with [1] and [2]. The results of these folds are summarized in Table 1.

Table 1 Performance comparison of complex human activity recognition time between three different methods

Number of sequences (S)	Complex Human Activity Recognition time –WEIZMANN dataset (%)			Complex Human Activity Recognition time –UIUC dataset (%)		
	GPC-SLSC (WEIZMANN dataset)	Compressive sensing dictionary based approach (WEIZMANN dataset)	LSTM Recurrent Neural Network (WEIZMANN dataset)	GPC-SLSC (UIUC dataset)	Compressive sensing dictionary-based approach (UIUC dataset)	LSTM Recurrent Neural Network (UIUC dataset)

10	0.15	0.22	0.23	0.24	0.28	0.38
20	0.2	0.22	0.28	0.25	0.27	0.35
30	0.22	0.25	0.35	0.27	0.3	0.38
40	0.25	0.3	0.4	0.3	0.35	0.4
50	0.2	0.28	0.32	0.22	0.25	0.3
60	0.23	0.27	0.25	0.25	0.27	0.33
70	0.3	0.33	0.38	0.32	0.34	0.37
80	0.35	0.41	0.45	0.4	0.45	0.47
90	0.27	0.35	0.4	0.3	0.32	0.35
100	0.3	0.31	0.35	0.32	0.35	0.4

Table 1 illustrates that the general performance of GPC-SLSC is the lowest. Since our algorithm can be applied for complex recognition of human activity, improvements on recognition time are obvious. Despite parameters being generalized on all two datasets, for the proposed and existing methods, results show that the GPC-SLSC framework still achieves better results than the state-of-the-art method. The sample calculation is provided below for two different datasets. Followed by which the graphical representation is given with detailed analysis.

Sample calculation (using WEIZMANN dataset)

- **Compressive sensing dictionary based approach**: With number of sequences being ‘10’, the time taken for complex human activity recognition for single video sequence was ‘$0.022ms$’, then the complex human activity recognition time for ‘10’ sequences are as given below.

$$HAR_t = 10 * 0.022ms = 0.22ms$$

- **LSTM Recurrent Neural Network**: With number of sequences being ‘10’, the time taken for complex human activity recognition for single video sequence was ‘$0.023ms$’, then the complex human activity recognition time for ‘10’ sequences is as given below.

$$HAR_t = 10 * 0.023ms = 0.23ms$$

- **Proposed GPC-SLSC**: With number of sequences being ‘10’, the time taken for complex human activity recognition for single video sequence was ‘$0.015ms$’, then the complex human activity recognition time for ‘10’ sequences is as given below.

$$HAR_t = 10 * 0.015ms = 0.15ms$$

Sample calculation (using UIUC dataset)

- **Compressive sensing dictionary based approach**: With number of sequences being '10', the time taken for human activity recognition for 1 sequence is '$0.028ms$', then the complex human activity recognition time for '10' sequences are as given below.

$$HAR_t = 10 * 0.028ms = 0.28ms$$

- **LSTM Recurrent Neural Network**: With number of sequences being '10', the time taken for human activity recognition for 1 sequence is '$0.038ms$', then the complex human activity recognition time for '10' sequences is as given below.

$$HAR_t = 10 * 0.038ms = 0.38ms$$

- **Proposed GPC-SLSC**: With number of sequences being '10', the time taken for human activity recognition for 1 sequence is '$0.024ms$', then the complex human activity recognition time for '10' sequences is as given below.

$$HAR_t = 10 * 0.024ms = 0.24ms$$

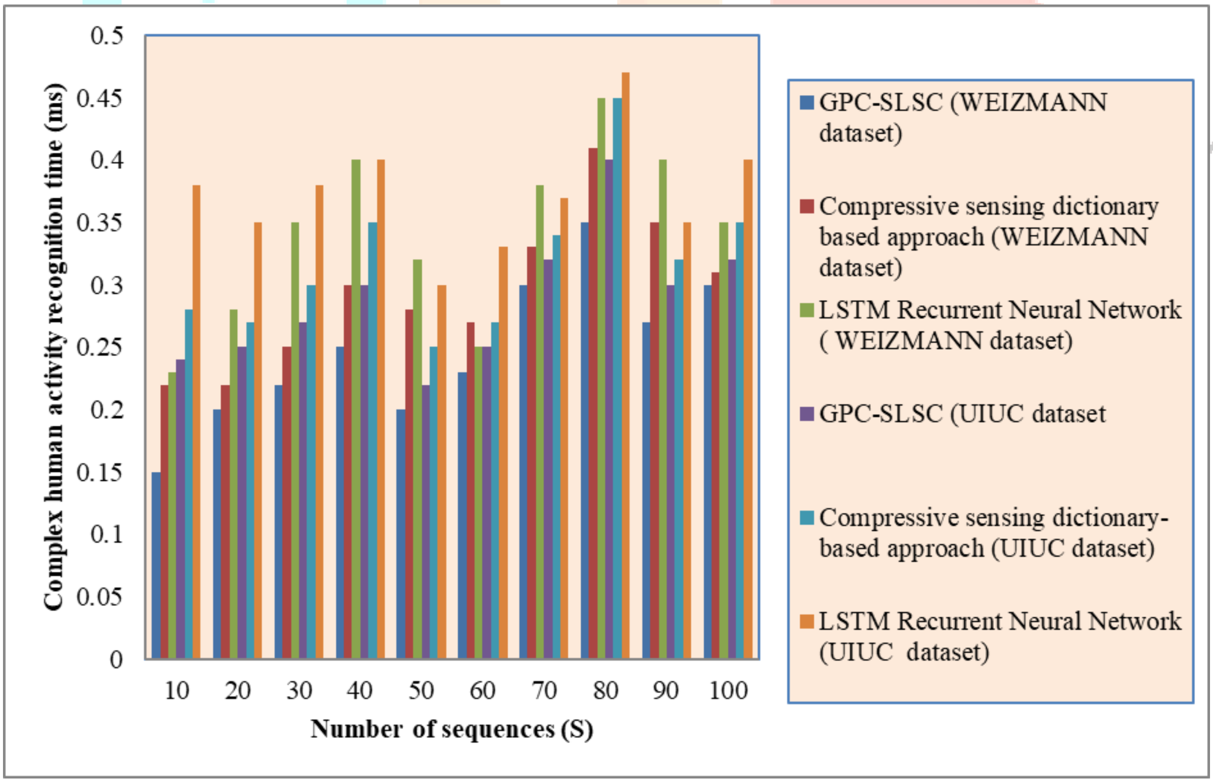

Figure 4 Performance comparison of complex human activity recognition time

Figure 4 given above shows the comparison performance of complex human activity recognition time for 10 different video sequences collected from 10 different persons involving various activities. As a result, 100 different sequences are observed in the x axis and complex human activity recognition time is observed in the y axis. With increase in the number of sequences using the two datasets the time taken for robust feature extraction also increases. As a result, complex human activity recognition time increases with the increase in the number of sequences. However, it is found to be comparatively better using GPC-SLSC when compared to Compressive sensing dictionary based approach [1] and LSTM Recurrent Neural Network [2]. This is because of the inherent foreground detection using the Gaussian Frame Detection algorithm. By applying Gaussian Frame Detection algorithm, differentiation between foreground and background are made in an efficient manner using Gaussian distribution that in turn extracts the actual frames. As a result, the time taken for recognition complex human activity is found to be comparatively lesser using GPC-SLSC by 16% compared to [1] and 27% compared to [2].

5.2 Scenario 2: Impact of Human activity recognition accuracy

In the second experiment, complex human activity recognition accuracy is performed and comparison made with [1] and [2]. The results of these folds are summarized in Table 2.

Table 2 Performance comparison of human activity recognition accuracy between three different methods

Number of sequences (S)	Human Activity Recognition Accuracy – WEIZMANN dataset (%)			Human Activity Recognition Accuracy (%) – (UIUC dataset)		
	GPC-SLSC (WEIZMANN dataset)	Compressive sensing dictionary-based approach (WEIZMANN dataset)	LSTM Recurrent Neural Network (WEIZMANN dataset)	GPC-SLSC (UIUC dataset)	Compressive sensing dictionary-based approach (UIUC dataset)	LSTM Recurrent Neural Network (UIUC dataset)
10	90	80	70	80	70	70
20	90	80	70	80	70	70
30	90	80	80	80	70	70
40	80	80	70	70	60	60
50	80	80	70	70	70	60
60	80	70	70	70	60	60
70	80	70	70	70	60	60
80	70	70	60	60	60	50
90	70	60	60	60	60	50

100	70	70	60	60	50	50

Table 2 summarizes results of the experiment to compare the human activity recognition accuracy performance between the proposed GPC-SLSC framework and the existing Compressive sensing dictionary-based approach [1] and LSTM Recurrent Neural Network [2]. The GPC-SLSC framework performs better than the two existing methods. The sample calculation is provided below for two different datasets with the graphical representation provided in figure 5.

Sample calculation (using WEIZMANN dataset)

- **Compressive sensing dictionary-based approach**: With '10' sequences given as input and successful recognition being '8', the complex human activity recognition accuracy is measured as given below.

$$HARA = \frac{8}{10} * 100 = 80\%$$

- **LSTM Recurrent Neural Network**: With '10' sequences given as input and successful recognition being '7', the complex human activity recognition accuracy is measured as given below.

$$HARA = \frac{7}{10} * 100 = 70\%$$

- **Proposed GPC-SLSC**: With '10' sequences given as input and successful recognition being '9', the complex human activity recognition accuracy is measured as given below.

$$HARA = \frac{9}{10} * 100 = 90\%$$

Sample calculation (using UIUC dataset)

- **Compressive sensing dictionary-based approach**: With '10' sequences given as input and successful recognition being '7', the complex human activity recognition accuracy is measured as given below.

$$HARA = \frac{7}{10} * 100 = 70\%$$

- **LSTM Recurrent Neural Network**: With '10' number of sequences given as input and successful recognition being '7', the complex human activity recognition accuracy is measured as given below.

$$HARA = \frac{7}{10} * 100 = 70\%$$

- **Proposed GPC-SLSC**: With '10' sequences given as input and successful recognition being '8', the complex human activity recognition accuracy is measured as given below.

$$HARA = \frac{8}{10} * 100 = 80\%$$

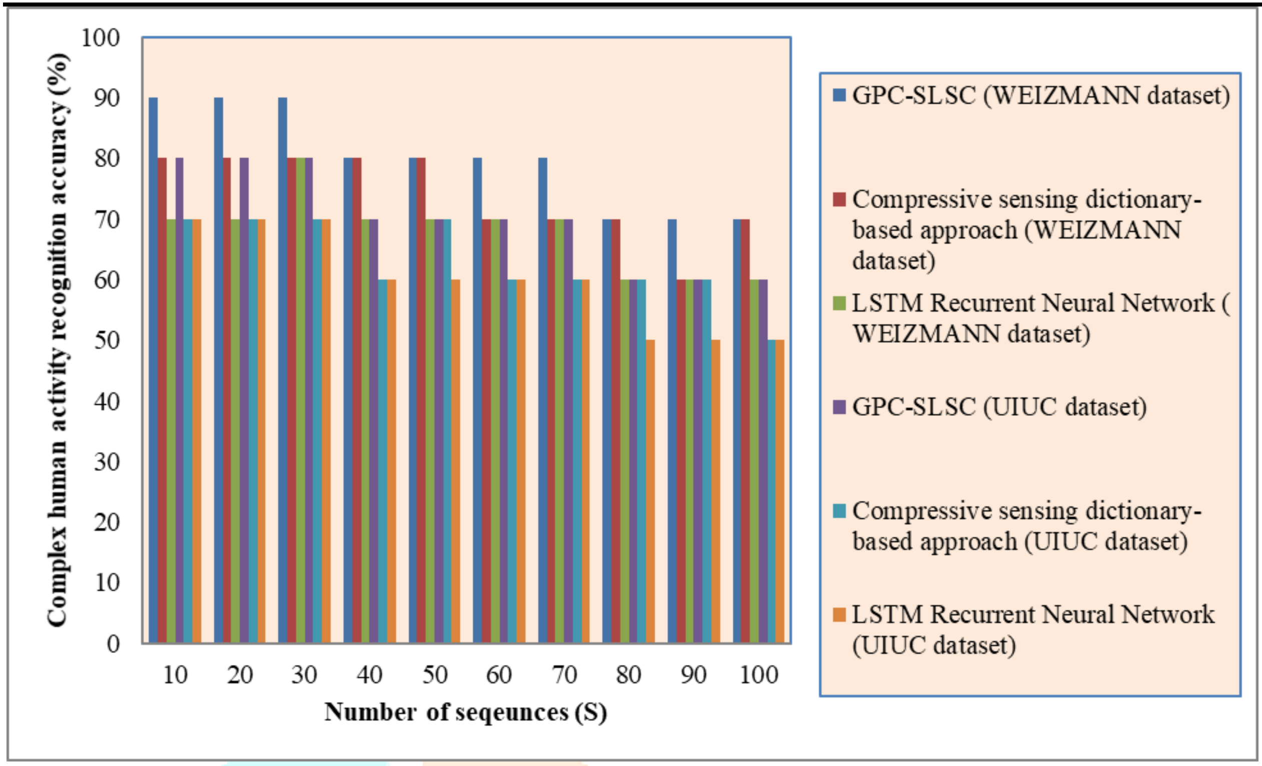

Figure 5 Performance comparison of complex human activity recognition accuracy

Figure 5 shows the performance comparison of complex human activity recognition accuracy for the WEIZMANN and UIUC datasets as the results of the experiment. In the figure, x axis represents the sequences and y axis represents the activity recognition accuracy. In the case of the UIUC dataset, the activity performed by eight actors with 532 high resolution sequences of 14 different activities showed the lowest recognition accuracy. This was because of different range of high resolution sequences considered for experimentation. For the WEIZMANN dataset, the recognition rate was said to be comparatively better than UIUC datasets due to similar resolution used. On the other hand, the activity recognition accuracy is inversely proportional to sequences, i.e., with the increase in the sequences, feature extraction time also increases because of varying resolution involved in different video sequences. As a result, the accuracy is also said to be reduced. However, with the application of Weighted Geometric Feature Extraction algorithm, center of mass for each video sequences were measured both along the x axis and y axis separately. Therefore, relevant features were extracted, therefore improving the recognition accuracy using GPC-SLSC by 11% when compared to [1] and 17% when compared to [2] using UIUC dataset.

5.3 Scenario 3: Impact of precision and recall

In the third experiment, precision and recall value is measured and comparison made with [1] and [2]. The results of these folds are summarized in Table 3 and table 4 respectively.

Table 3 Performance comparison of precision over three different methods

Number of sequences	Precision (%) –WEIZMANN dataset			Precision (%) – (UIUC dataset)		
	GPC-SLSC (WEIZMANN	Compressive sensing	LSTM Recurrent	GPC-SLSC	Compressive sensing	LSTM Recurrent

(S)	dataset)	dictionary-based approach (WEIZMANN dataset)	Neural Network (WEIZMANN dataset)	(UIUC dataset)	dictionary-based approach (UIUC dataset)	Neural Network (UIUC dataset)
10	94.32	88.23	85.13	95.13	82.13	75.56
20	90.15	86.14	82.23	92.35	85.13	80.13
30	88.32	82.13	80.15	90.45	88.14	82.14
40	85.12	81.32	78.95	88.15	80.23	75.89
50	89.23	85.13	82.13	93.25	84.13	80.14
60	90.14	80.23	75.23	92.14	85.56	80.13
70	92.35	78.14	79.13	95.56	87.23	82.13
80	85.14	75.56	72.13	90.13	83.13	79.25
90	83.12	78.25	76.13	85.13	84.22	80.14
100	81.25	75	68.75	88.88	88.23	87.5

Table 4 Performance comparison of recall over three different methods

Number of sequences (S)	Recall (%) –WEIZMANN dataset			Recall (%) – (UIUC dataset)		
	GPC-SLSC (WEIZMANN dataset)	Compressive sensing dictionary-based approach (WEIZMANN dataset)	LSTM Recurrent Neural Network (WEIZMANN dataset)	GPC-SLSC (UIUC dataset)	Compressive sensing dictionary-based approach (UIUC dataset)	LSTM Recurrent Neural Network (UIUC dataset)
10	89.23	85.13	80.14	93.14	88.25	82.14
20	91.35	88.13	82.32	94.32	90.14	85.13
30	94.23	84.23	80.25	96.77	93.25	90.24
40	85.14	80.25	75.13	90.25	88.45	80.14
50	89.25	81.32	70.33	91.32	87.14	83.25
60	90.14	87.23	81.24	92.48	89.25	87.89
70	86.23	80.25	75.78	90.15	85.13	80.14
80	82.14	78.13	73.22	85.22	80.14	79.13
90	78.23	70.13	78.14	80.24	78.32	77.15
100	65	60	55	80	75	70

Table 3 and table 4 summarize the results of the experiment to compare the precision and recall performance among the proposed GPC-SLSC framework and two state-of-the-art methods. In this experiment, the proposed GPC-SLSC framework performed better than the two state-of-the-art methods for both the WEIZMANN and the UIUC datasets. This result implies that Sequence Likelihood Support Vector Machine Classifier is very effective to learn the complex human activities. The sample calculation is provided below for two different datasets with the graphical representation provided in figure 6 and figure 7.

Sample calculation: Precision and recall (using WEIZMANN) for GPC-SLSC with 100 sequences

A dataset containing 100 video sequences are considered for experimentation for activity recognition using WEIZMANN dataset. A search was conducted for activity recognition and 80 sequences were retrieved. Of the 80 sequences retrieved, 65 were relevant. Using the designations above, let A represents the number of relevant sequences retrieved, let B represents the number of relevant sequences not retrieved, and C represents the number of irrelevant sequences retrieved. Then, A = 65, B = 35 (100-65) and C = 15 (80-65). Then, precision and recall using WEIZMANN for GPC-SLSC with 100 sequences is given below.

$$Recall = 65/(65 + 35) * 100 = 65\%$$

$$Precision = 65 / (65 + 15) * 100 = 81.25\%$$

Precision and recall (using WEIZMANN) for Compressive sensing dictionary-based approach with 100 sequences

$$Recall = 60/(60 + 40) * 100 = 60\%$$

$$Precision = 60 / (60 + 20) * 100 = 75\%$$

Precision and recall (using WEIZMANN) for LSTM Recurrent Neural Network with 100 sequences

$$Recall = 55/(55 + 45) * 100 = 55\%$$

$$Precision = 55 / (55 + 25) * 100 = 68.75\%$$

Sample calculation: Precision and recall (using UIUC) for proposed GPC-SLSC with 100 sequences

A dataset containing 100 sequences are considered for experimentation for activity recognition using UIUC dataset. A search was conducted for activity recognition and 90 sequences were retrieved. Of the 90 sequences retrieved, 80 were relevant. Using the designations above, let A represents the number of relevant sequences retrieved, let B represents the number of relevant sequences not retrieved, and C represents the number of irrelevant sequences retrieved. Then, A = 80, B = 20 (100-80) and C = 10 (90-80), then, precision and recall using UIUC for GPC-SLSC with 100 sequences is given below.

$$Recall = 80/(80 + 20) * 100 = 80\%$$

$$Precision = 80 / (80 + 10) * 100 = 88.88\%$$

Precision and recall (using UIUC) for Compressive sensing dictionary-based approach with 100 sequences

$Recall = 75/(75 + 25) * 100 = 75\%$

$Precision = 75/(75 + 10) * 100 = 88.23\%$

Precision and recall (using UIUC) for LSTM Recurrent Neural Network with 100 sequences

$Recall = 70/(70 + 30) * 100 = 70\%$

$Precision = 70/(70 + 10) * 100 = 87.5\%$

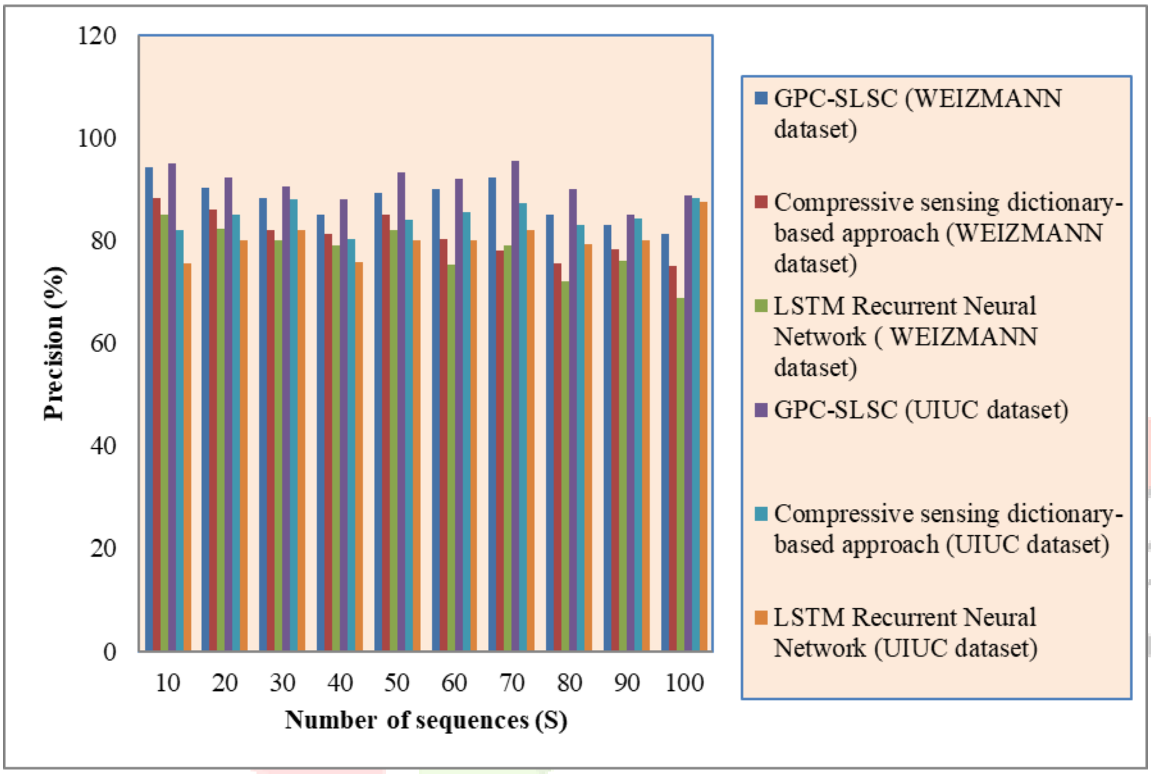

Figure 6 Performance comparison of precision

Figure 6 given above shows the results of the experiment to compare the precision rate performance between the proposed GPC-SLSC and existing Compressive sensing dictionary-based approach [1], LSTM Recurrent Neural Network [2] respectively. The proposed GPC-SLSC performs better than the state-of-the-art works. The performance of GPC-SLSC framework with respect to precision made a significant improvement when the sequence size was less, and the rate of precision fell down upon increase in the sequence size. Hence, it indicates that incorporating short sequences were found to be efficient. This is because with the increase in the sequences, the activity to be monitored and recognized gets increases and therefore a decrease is said to be observed with minimum sequence size and vice versa. Figure 6 show that the GPC-SLSC framework achieves remarkable improvements using both the datasets. The improvement attributes to their maximum likelihood representation and probability sequences which is preferred by the GPC-SLSC framework. Results also show that the precision rate, which is generated by implementing Sequence Likelihood Support Vector Machine Classifier algorithm on the video sequences, obtains the highest performance within group comparison on all two datasets. Hence, the precision rate using GPC-SLSC framework is found to be improved by 8% compared to Compressive sensing dictionary-based approach [1] and 14% compared to LSTM Recurrent Neural Network [2] respectively.

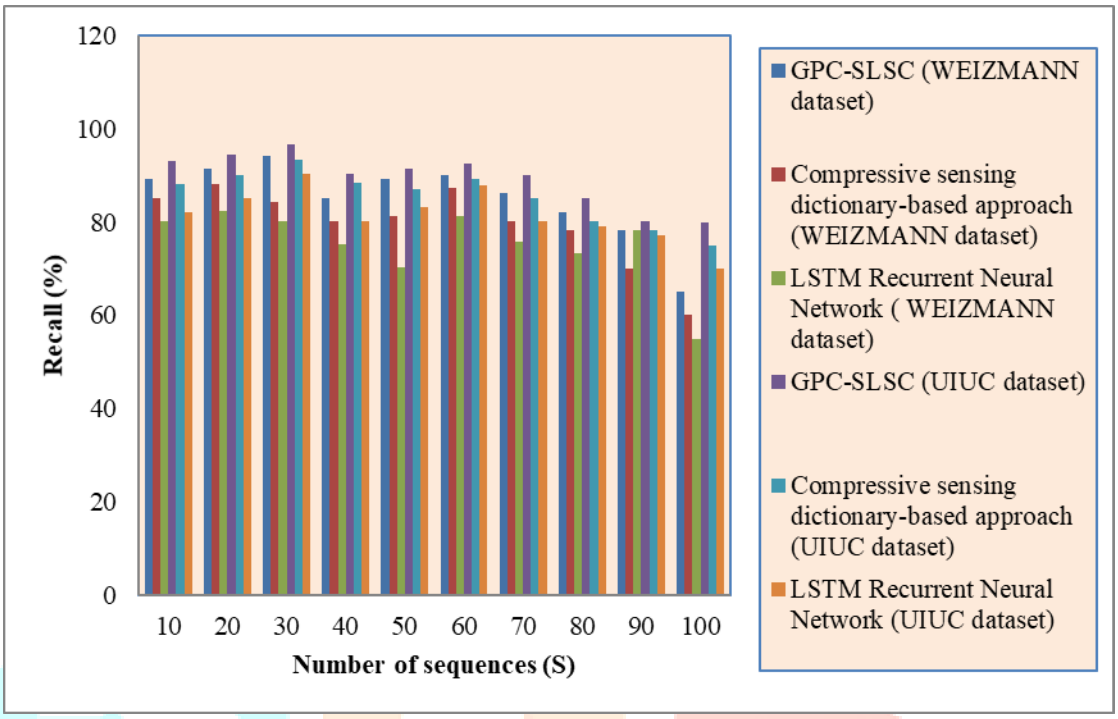

Figure 7 Performance comparison of recall

Figure 7 shows that GPC-SLSC framework performs relatively better than for Compressive sensing dictionary-based approach [1] and LSTM Recurrent Neural Network [2] with respect to recall rate. As the Compressive sensing dictionary-based approach focus on atomic postures, it contains information from a single and simple perspective. By contrast, the GPC-SLSC framework, extracts robust features and using polar coordinates that serves for complex recognition. Besides, maximum likelihood with dual coordinate estimator shows advantages in this condition. This is because in the modeling process, polar coordinates based on the center of mass for each sequence have inherent characters for feature extraction. This property is especially deployed in GPC-SLSC framework that classifies with maximum likelihood of probability sequences. All the factors mentioned above contribute to gaining better performances with respect to recall rate when deploying GPC-SLSC framework. The performance improvement of recall rate using GPC-SLSC framework is improved by 7% compared to Compressive sensing dictionary-based approach [1] and 14% compared to LSTM Recurrent Neural Network [2] using WEIZMANN dataset.

6. Conclusion

In this paper, we present the Gaussian Polar Coordinate with Sequence Likelihood Support Classifier (GPC-SLSC) framework for recognition of complex human activity. The framework consists of three parts, frame detection, robust feature extraction and complex human activity recognition. Gaussian Foreground Frame Detection is applied to the input dataset to extract the inherent foreground features. Next, Polar Coordinate Robust Feature Extraction model is applied to the inherent foreground features for robust extraction of features. Finally, Maximum Sequence Likelihood Support Vector Machine Classifier is applied to the extracted features for recognizing complex human activity. The GPC-SLSC has been implemented to validate the proposed framework. Extensive experiments conducted using two datasets demonstrate the effectiveness and robustness of our proposed framework in terms of computational time, complex human activity recognition accuracy, precision and recall.

References

[1] Lina Yao, Quan Z. Sheng, Xue Li, Tao Gu, Mingkui Tan, Xianzhi Wang, Sen Wang, and Wenjie Ruan, “Compressive Representation for Device-Free Activity Recognition with Passive RFID Signal Strength”, IEEE Transactions on Mobile Computing (Volume: 17, Issue: 2, Feb. 1 2018)

[2] Deepika Singh, Erinc Merdivan, Ismini Psychoula, Johannes Kropf, Sten Hanke, Matthieu Geist and Andreas Holzinger, “Human Activity Recognition Using Recurrent Neural Networks”, International Federal for Information Processing, Springer, May 2017

[3] S. Gaglio, G. Lo Re, M. Morana, “Human Activity Recognition Process Using 3-D Posture Data”, IEEE Transactions on Human-Machine Systems, May 2014

[4] Mohammed Mehedi Hassan, Md. Zia Uddin, Amr Mohamed, Ahmad Almogren, “A robust human activity recognition system using smartphone sensors and deep learning”, Future Generation Computer Systems, Elsevier, Nov 2017

[5] Amir H. Shabani, John S. Zelek, David A. Clausi, “Multiple scale-specific representations for improved human action Recognition”, Pattern Recognition Letters, Elsevier, Jan 2013

[6] J. Suto, S. Oniga, P. Pop Sitar, “Feature Analysis to Human Activity Recognition”, International Journal of Computers Communications & Control, Feb 2017

[7] Xiaoxia Huang, Mingwei Dai, “Indoor Device-Free Activity Recognition Based on Radio Signal”, IEEE Transactions on Vehicular Technology (Volume: 66, Issue: 6, June 2017)

[8] Md Ferdous Wahid, Reza Tafreshi, Mubarak Al-Sowaidi, Reza Langari, “Subject-Independent Hand Gesture Recognition using Normalization and Machine Learning Algorithms”, Journal of Computational Science, Elsevier, Apr 2018

[9] Daniele Rav, Charence Wong, Benny Lo, and Guang-Zhong Yang, “A Deep Learning Approach to on-Node Sensor Data Analytics for Mobile or Wearable Devices”, IEEE Journal of Biomedical and Health Informatics, VOL 21, NO 1, JANUARY 2017

[10] Sofia Savvaki, Grigorios Tsagkatakis, Athanasia Panousopoulou, and Panagiotis Tsakalides, “Matrix and Tensor Completion on a Human Activity Recognition Framework”, IEEE Journal of Biomedical and Health Informatics, VOL 21, NO 6, NOVEMBER 2017

[11] Sara Khalifa, Guohao Lan, Mahbub Hassan, Aruna Seneviratne, and Sajal K. Das, “HARKE: Human Activity Recognition from Kinetic Energy Harvesting Data in Wearable Devices”, IEEE Transactions on Mobile Computing (Volume: 17, Issue: 6, June 1 2018)

[12] Lei Wang, Xu Zhao, Yunfei Si, Liangliang Cao, and Yuncai Liu, “Context-associative Hierarchical Memory Model for Human Activity Recognition and Prediction”, IEEE Transactions on Multimedia (Volume: 19, Issue: 3, March 2017)

[13] Alexandros Andre Chaaraoui, Pau Climent-Perez, Francisco Florez-Revuelta, "Silhouette-based human action recognition using sequences of key poses", Pattern Recognition Letters, Elsevier, Feb 2013

[14] Manuel J. Marin-Jimenez, Enrique Yeguas, Nicolas Perez de la Blanca, "Exploring STIP-based models for recognizing human interactions in TV videos", Pattern Recognition Letters, Elsevier, Nov 2012

[15] Alexandros Iosifidis, Anastasios Tefas, Ioannis Pitas, "Dynamic action recognition based on dynemes and Extreme Learning Machine", Pattern Recognition Letters, Elsevier, Nov 2012

[16] Xiantong Zhen, Ling Shao, "A local descriptor based on Laplacian pyramid coding for action Recognition", Pattern Recognition Letters, Elsevier, Nov 2012

[17] M.M. Youssef, V.K. Asari, "Human action recognition using hull convexity defect features with multi-modality setups", Pattern Recognition Letters, Feb 2013

[18] Salvatore Gaglio, Giuseppe Lo Re, and Marco Morana, "Human Activity Recognition Process Using 3-D Posture Data", IEEE Transactions on Human-Machine Systems (Volume: 45, Issue: 5, Oct. 2015)

[19] Shugang Zhang, Zhiqiang Wei, Jie Nie, Lei Huang, Shuang Wang, and Zhen Li, "A Review on Human Activity Recognition Using Vision-Based Method", Hindawi Journal of Healthcare Engineering Volume 2017

[20] XIA Li-min, HAN Fen, WANG Jun, "Complex human activities recognition using interval temporal syntactic model", Springer, Dec 2015

[21] J. Gall and V. Lempitsky, "Class-specific hough forests for object detection," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR '09), pp. 1022–1029, June 2009

[22] D. Tran and A. Sorokin, "Human activity recognition with metric learning," in Proceedings of the 10th European Conference on Computer Vision (ECCV '08), pp. 548–561, 2008