International Journal on Future Revolution in Computer Science & Communication Engineering ISSN: 2454-4248
Volume: 3 Issue: 11 462 – 471

A Survey on Human Activity Analysis Techniques

Bhoomika Rathod(PG student)[1*], Dr. Devang Pandya [2], Prof. Raunakraj Patel [3]
[*]Information Technology, Silver Oak College of Engineering and Technology,
Gujarat Technological University, Ahemdabad, *India*
[1]brathod446@gmail.com, [2] devang1979@gmail.com,[3]raunakrajpatel.it@socet.edu.in

Abstract— Human Activity Recognition(HAR) is Popular research topic in Computer vision and Image Processing area. This Paper Provide an exhaustive survey on the Entire Process of identify or Recognize Human activity. Basically, There are Four steps are involved in HAR process, which are Pre-processing, Feature extraction, Training, and Classification of different activities from video. The need of data preprocessing , and segmentation based on camera movements are presented. This paper provide detailed survey on different features for HAR, feature extraction and selection method , and Classification methods with advantages and disadvantages. Finally, A brief discussion about various classification techniques are presented.

Keywords*- HAR process, Static & Moving camera segmentation, Feature extraction, Feature selection, Classification & Recognition.*

*****

I. INTRODUCTION

Recognition of Human activity is a complex, diverse, and challenging area that has received significant research interest within few past years. The Primary goal of HAR is to accurately detect human activities in real-life. Recognition of Human activity from video has drawn attention because of its growing need in various real life environment such as Security Surveillance for effective monitoring of public places such as airports, railways stations, shopping malls, Crowded sport arenas, military installation etc..,or for smart healthcare facilities such as daily activity monitoring and fall detection in old people's homes. Human activity depends on the context and environment where it is being performed . It has different levels of complexity such as Body Parts motion(gestures & sign languages),Movements(walking, running), Actions(reaching & following), Human-Object interaction(grasping & punching), Human-Human interaction(handshaking & punching), and Social behavior (leading & emphazing) [1].

In general, The Recognition of Human activity from video involves four different steps, i.Acquisition of input video and Pre-Processing (extraction of image or frames, Filtering for remove noise etc..,segmentation of ROI). ii. Feature extraction is performed by Motion tracking ,in which basic idea is to detect moving object in frame & extract the features like Data of pose, joint point of skeleton, blob trajectory, histogram etc... iii. Training given to algorithm using Feature vector obtained from different features to classify different movements. This algorithm referred to as "Classifier". iv. Finally, Classification is performed by classifier to identify & investigate the activities from the videos.

This paper present the detailed Survey on entire process of HAR and involved diff. techniques. The rest of the paper is organized as follows: section II briefly presents a review of related works, section III present the block diagram of HAR process, section VI describe the Input video and Data-Preprocessing, section V describe Segmentation and its type briefly, section VI present the various Feature extraction and selection method , section VII present the Classification of Human activities in detail.

II. RELATED WORK

Kamrad Khosal Roudposhti et al. [1] explains how the body motion based features can be analyzed, They propose a multilevel framework able to deal with the different levels of human activities. Framework represent dependency between different level of Human activity analysis. Using Motion track suit, acceleration signal and 3D position Data of the Human body part is obtained , which are used to LLF(low level feature) extraction. LMA(Laban movement analysis) is used to describe LLF like body parts frequency, ΔH(variation in height of body) as a descriptor. LMA uses Effort, Shape, Relationship components to describe LLF, and the Multi-level activity of person is estimated using this components quality. Each level is depend on previous level output, and the entire method is implemented using Bayesian network. There are no need to train every level of Framework each time, only New level is needed to be trained.

B.Ay [2] proposed an approach for efficient human activity recognition using relations between motion data taken joint data positions from skeleton sequences. they proposed a feature extraction method choosing key joints of a skeleton model and creating a biomechanical model that depends on motion changes. key joints related to the motion determined by analysing image frames with data of joints positions obtained from a kinect sensor camera. The key joints with the biomechanical model will facility recognizing motion better. First of all Motion database is collected using Biomechanical model by capturing the position data of human using kinect camera. which give the skeleton of human body with 3D data of joint points. These 3d data is used to feature extraction. In feature extraction , key joint points are selected using position and velocity data. At last, Key joint match algorithm is used to implement entire method , and recognize Human activity like lifting, sitting down and hand waving.

Chen Hong et al. [3] present novel real-time motion recognition based on skeleton animation which is capable of

 ISSN: 2454-4248
462 – 471

capturing and recognizing motion with high sensitivity and precision and also addressing other objectives such as real-time, robustness, and extensibility. First inertial based Motion capture is performed which give the real-time & reliable skeleton motion data as input for motion recognition.That contain the 18 inertial sense units. A Five level Hierarchical skeleton model is constructed for feature extraction, in which First head is chosen as a root. Second, the torso nodes include spine1,spine2, left shoulder, right shoulder and hips. Third, the limb level 1 nodes include left arm, right arm, left up leg and right up leg. Fourth, the limb level 2 nodes include left fore-arm, right fore-arm, left leg and right leg. Fifth, the limb level 3 nodes include left hand, right hand, left foot and right foot. By giving the training the classifier , a set of strong classifiers is obtained used to Motion recognition.At the last, HMM is used to recognize motion which work as a smoothing filter that reduce the noise and smooth the sequence of recognized class.

A detailed survey on HAR shows that the demand for intelligent video processing is getting greater and greater, moving body behavior classification from video images. Congjie Zou et al. [4] deal with the Key challenge of automatically estimate the behaviour of a person or a body part from Multi-view video image. The first step is, Modelling of human by skeleton, which describe motion joints well. Then Establish a coordinate system with origin abdomen joint using the motion joints. For feature extraction , the Angle made by the each joints with vertical, which make each joints vertical for determine the postures of a moving human clearly, length-width ratio is added in feature. At the last, Bayesian networks implement method for motion recognition. The speed and accuracy has been improved greatly compared to the existing algorithms.

Serban Oprisescu et al. [5] deal with the issue of action recognition as an application of the new 3D time-of-flight camera, exploiting the special ability of the device to measure distance. ToF camera one can develop applications where even submilimeter movements can be sensed, the distance error could be high(0.5m) depending on the scene as well as on the camera optics and geometry. They focused on finding the way to correct these distance errors and the next generation of ToF cameras was deliver correct distance images and so extraction of the object of interest from image/video will become a simple task. In first step of the method, Segmentation or silhouette extraction from each image of video sequence using camera, which track the key joint of the human. The second step is to simplify as much as possible the "space-time shape" obtained by temporal concatenation of 2D silhouettes extracted in the segmentation phase. The result of the second step is replacing the space-time shape with the trajectories of the key points. The next step in simplifying the description of the action is to replace the trajectories with some features of them. Features are calculated using the function of variation, Total variation, real mean speed, and absolute mean speed of key points which give the information about key point position of continues frames of video. At the last, Decision tree is used for implement method and recognize action.

III. BLOCK DIAGRAM OF HAR PROCESS

The "Fig. 1", shows the block diagram of the human activity recognition. The Recognition of Human activity from video involves four different steps, Acquisition of input video and Data Pre-Processing. Feature extraction is performed by Motion tracking ,in which basic idea is to detect moving object in frame & extract the features. Training given to algorithm using Feature vector obtained from different features to classify different movements. This algorithm referred to as "Classifier". Classification and Recognition is performed by classifier to identify & investigate the activities from the videos.

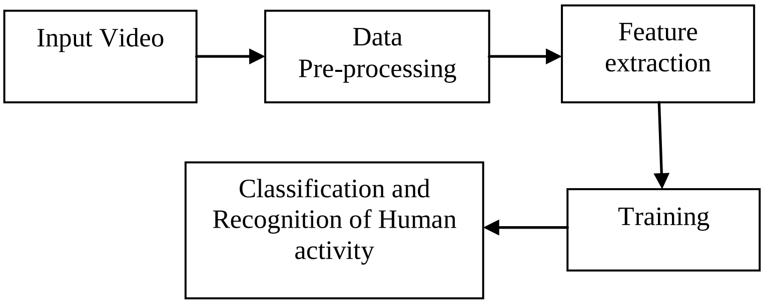

Fig. 1 Block Diagram of HAR Process.

IV. INPUT VIDEO & DATA PREPROCESSING

A recognition system obtains the environment information from visual inputs such as cameras. These inputs can be either a still image or a sequence of images [6]. These images or Video used as a Training video, comprises the experience that the algorithm uses to learn [7]. A Quality of data affects the used methods, raw data is pre-processed to improve the quality of data [2]. Data Pre-Processing is Primary step which transforms data into a format that will be more easily and effectively processed. It describes any type of processing performed on raw data to prepare it for another procedure. It capture the data to decrease the variation that causes a reduction in the recognition rate and increases the complexities. some techniques of Data pre-processing Image enhancement(filtering), Noise removal, Skew detection/correlation, Segmentation(Segmentation step aims to identify moving objects), Size normalization, Morphological processing(erosion & dilation, opening & closing, outlining, and thinning & skeletonisation) [8].

V. SEGMENTATION

The processed image can be used toward the segmentation step. A few systems have this part [9]. In a Video analysis application , Where analysis of human activities is carried out , the basic step is the detection of human movements. Detection of human Movements comprise Segmentation as a basic step , which is the process of dividing the image into dissimilar portion so that all together these segments results in the Original image [10]. Segmentation aims to identify moving object. In video segmenting moving objects are realized as a difficult problem because in addition to the variation of moving objects in the image, the camera moves as well [11]. According to camera movements, the segmentation are classified as follow:

A. Static Camera Segmentation

The Camera is placed in a specific location in a certain angle. so the angle of view for Object and the background is fixed. The usual way of human or object detection in this case is subtraction of the background [12]. The idea of background subtraction is to subtract or difference the current image from a reference background identifies non- stationary or new objects [13]. The simple background model assumes that intensity value of a pixel can be modeled by a single unimodel distribution. This basic model can't handle multiple backgrounds [14] [15]. The generalized mixture of Gaussian (MoG) in [16] has been used to model complex, non-static background.

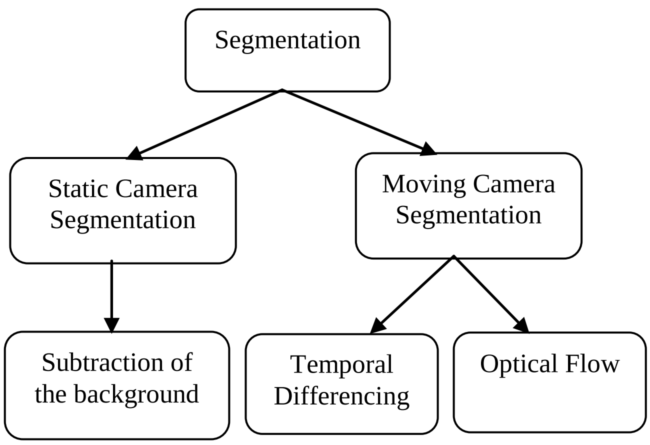

Fig. 2 Classification of Segmentation.

B. Moving Camera Segmentation

There are enormous challenges in moving camera segmentation. Because in addition to the movement of the object , It should consider the Camera movements and the background. In this case the temporal differencing [17] and optical flow [18] methods are used. Temporal differencing detects the change by subtracting Two or three consecutive frame and threshold. but temporal differencing has a major disadvantage that it only detects the boundary of the objects which requires to be filled by some post processing methods [19]. Optical flow detects motion within each neighborhood of pixels by registering changes in color & intensity of pixels form frame to frame. Optical flow classified as a dense based and point based optical flow [20].

VI. Feature Extraction & Feature Selection

Many machine learning application require feature extraction and feature selection.

A. Feature Extraction

Feature extraction can be seen as a pre-processing step in learning process where different kinds of features will be extracted from data[21]. The feature extraction step is possibly the most important part of the activity recognition problem since classification can be handled by any existing machine learning algorithm if the features are robust [22]. In the first step the data will be split into short intervals windows. Usually, a Window covers one or two seconds long time interval & its size depends on the Sampling frequency. After the windowing step, features will be extracted from each window[21].Feature extraction transforming the existing features into a lower dimensional space. It Creates a subset of new features by combinations of the existing features [22].
While numerous features can be extracted from physical activity signals, increasing the number of features does not necessarily increase the classification accuracy since the features may be redundant or may not be Class-specific. Features can be classify based on three class [23] :

(i) Time domain features: These features are typically used in many practical HAR systems because of being less computationally intensive; thus, they can be easily extracted in real time.

(ii) Frequency domain features: domain features require higher computational cost to distinguish between different human activities.

(iii) Physical features: are derived from a fundamental understanding of how a certain human movement would produce a specific sensor signal. Physical features are usually extracted from multiple sensor axes, based on the physical parameters of human movements.

Following, Table 1. present the Some Feature extraction methods proposed for HAR [21] [23].

TABLE I
SOME FEATURE EXTRACTION METHODS PROPOSED FOR HAR

Category	Features
Time-Domain	Mean, Variance, Mean Absolute Deviation, Root mean square, Zero Crossing Rate, Interquartile Range, 75'th percentile, Kurtosis, Signal magnitude area, Min-Max
Frequency-Domain	Spectral energy, Spectral entropy, Spectral centroid, Principal Frequency
Physical features	Object or blob trajectory(position, velocity, size, centroid, etc...)
	Histogram of (Motion, direction, color, texture, pixel change, n-gram events)
	spatial velocity & curvature
	Spatio-temporal features(patches, saliency)
	shape(bounding box, silhouette, image contours)
	optical flow(blob boarder, pixel level, spatio-temporal volume of force flow)
	R-Transform of silhouette shape
	Eigen motion
	2D Contour angles
	Visual words(location, motion, size)
	3D Gaussian distribution of spatio-temporal gradients

ISSN: 2454-4248
462 – 471

B. Feature selection

Feature selection, also called feature reduction , is the process of choosing a subset of original features according to a well-defined evaluation criterion. It is a frequently used dimensionality reduction techniques which removes irrelevant and redundant features. This approach has more useful effects for real applications because it accelerates algorithms improves the performance & simplifies the model [21]. Successful selection of features allows for simplification of models to make them easier to interpret, to decrease model training times, and to better understand difference between class [23]. Feature selection choosing a subset of all the features.

There are four basic steps in a typical features selection method showing as follows [24] :

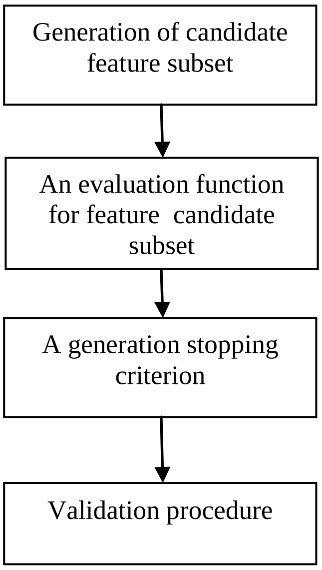

Fig. 3 Steps of Feature selection method.

Depending on the feature evaluation process, feature selection algorithms belongs to three different groups [21] :

(i) Filters: Filters calculate Scores for all features and select features according to the scores. It extract features from the data without any learning involved. Filters work without taking the classifier into consideration This makes them very computationally efficient. They are divided into multivariate and univariate methods. Multivariate methods are able to find relationships among the features, while univariate methods consider each feature separately .Gene ranking is a popular statistical method [25].

(ii) Wrappers: Wrappers tend to perform better in selecting features since they take the model hypothesis into account by training and testing in the feature space. This leads to the big disadvantage of wrappers, the computational inefficiency which is more apparent as the feature space grows. Unlike filters, they can detect feature dependencies. Wrappers are separated in 2 categories: Randomised and Deterministic.

(iii) Embedded modules: Feature selection takes place at the training process. which combine the feature selection step and the classifier construction[21]. Embedded techniques tend to do better computationally than wrappers but they make classifier dependent selections that might not work with any other classifier. That is because the optimal set of genesis built when the classifier is constructed and the selection is affected by the hypotheses the classifier makes. A well-known embedded technique is random forests[25]. A random forest is a collection of classifiers. New random forests are created iteratively by discarding a small fraction of genes that have the lowest importance [26].

Some of the Feature Selection method applied on HAR

(1) Principal Component Analysis (PCA), Independent component analysis (ICA) & Linear discriminant analysis (LDA): PCA is an unsupervised second order statistical approach to find useful basis for data representation. It finds PCs at the optimally reduced dimension of the input. For human activity recognition, it focuses on the global information of the binary silhouettes, which has been actively applied. However, PCA is only limited to second order statistical analysis, allowing up to decor relation of data. The role of PCA is to approximate the original data with lower dimensional features. Its fundamental is to compute the eigenvectors of the covariance data matrix and then the approximation is done using a linear combination of a few top eigenvectors [27]. ICA is a computational method for separating a multivariate signal into additive subcomponents supposing the mutual statistical independence of the non-Gaussian source signals. It is a special case of blind source separation. It finds the independent components by maximizing the statistical independence of the estimated components. ICA can be define by "Minimization of Mutual Information" and "Maximization of non-Gaussianit"[28]. In [27] local binary silhouette features through ICA to represent human body in different activities usefully is discussed. However, as PCA considers the second order moments only, it lacks information on higher order statistics. On the contrary, ICA considers higher order statistics and it identifies the independent source components from their linear mixtures. Hence, ICA provides a more powerful data representation than PCA as it tries to provide an independent rather than uncorrelated feature representation. LDA is an efficient classification tool that works based on grouping of similar classes of data. The LDA algorithm seeks the vectors in the underlying space to create the best discrimination among different classes. It finds the directions along which the classes are best separated by considering the within- class scatter but also the between-class scatter [29].

(2) Minimum Redundancy Maximum Relevance (mRMR): MRMR is a method that maximises the relevancy of genes with the class label while it minimises the redundancy in each class. To do so, it uses several statistical measures. Mutual Information (MI) measures the information a random variable can give about another, in particular the gene activity and the class label. The method can be applied to both categorical and continuous variables. The method can be applied to both categorical and continuous variables [25]. For categorical (discrete) variables, MI(Mutual information) is used to find genes that are not redundant (minimize redundancy) W and are maximally relevant V with a target label[30]. For continuous variables the F-statistic is used to

find the maximum relevance between a gene and a class label and then the correlation of the gene pair in that class is measured to minimise redundancy[30]. MR methods give lower error accuracies for both categorical and discrete data.

(3) Correlation-based feature selection (CFS): CFS follows he principal that "a good feature subset is one that contains features highly correlated with the class yet uncorrelated with each other." [31]. CFS evaluates a subset by considering the predictive ability of each one of its features individually and also their degree of redundancy (or correlation). The difference between CFS and other methods is that it provides a "heuristic merit" for a feature subset instead of each feature independently [32]. This means that given a function (heuristic), the algorithm can decide on its next moves by selecting the option that maximizes the output of this function. Heuristic functions can also be designed to minimize the cost to the goal.

(4) Fast Correlation-based filter (FCFB): FCBF is a multivariate feature selection method which starts with full set of features, uses symmetrical uncertainty to calculate dependences of features and finds best subset using backward selection technique with sequential search strategy. It has an inside stopping criterion that makes it stop when there are no features left to eliminate. It is a correlation based feature subset selection method which runs, in general, significantly faster than other subset selection methods [33]. Previous experiments show that FCBF is an efficient and fast algorithm which uses interdependence of features together with the dependence to the class. It selects best subset of features from the full set by means of backward elimination. Especially when inputs are highly correlated, this method may eliminate too many features. Baris Senliol et al.[34] introduce a new approach where we change the elimination method and the new algorithm is called FCBF#. FCBF# changes FCBF's quick and sharp elimination method to a more balanced one to select the best subset which has ***K*** features.

(5) Rankfeatures: Rankfeatures ranks features by a given class separability criterion. Class separability measures include the absolute value of a statistic of a two-sample t-test, Kullback-Leibler distance, minimum attainable classification error, area between the empirical Receiver Operating Characteristic (ROC) curve and the random classifier slope, and the absolute value of the statistic of a two-sample unpaired Wilcoxon test. Measures are based on distributional characteristics of classes (e.g., mean, variance) for a feature [23].

(6) ReliefF: ReliefF [33] is a commonly used filter method that ranks features by weighting them based on their relevance. Feature relevance is based on how well data instances are separated. For each data instance, the algorithm finds the nearest data point from the same class (hit) and nearest data points from different classes (misses).

(7) SIFT & SURF: Scale-Invariant Feature Transform (SIFT) is a method for extracting features from an image that is less sensitive to the scale, rotation, and illumination of the image. Each SIFT feature, or descriptor, is a vector that describes edges and corners in a region of an image. SIFT also captures information about the composition of each point of interest and its surroundings. Speeded-Up Robust Features (SURF) is another method of extracting interesting points of an image and creating descriptions that are invariant of the image's scale, orientation, and illumination. SURF can be computed more quickly than SIFT, and it is more effective at recognizing features across images that have been transformed in certain ways [34].

VII. Classification & Recognition

HAR includes Machine learning Systems are often described as learning from experience either with or without supervision from humans. The classifiers need to learn the patterns in feature data before they can recognize the patterns associated to activities.

(i) Supervised learning problem: a program predicts an output for an input by learning from pairs of labeled inputs and outputs; that is, the program learns from examples of the right answers. For example, assume that you have collected data describing the heights and weights of people. Now assume that the data is also labeled with the person's sex. An example of as supervised learning problem is inducing a rule to predict whether a person is male or female based on his or her height and weight [34].

(ii) Unsupervised learning problem: a program does not learn from labeled data. Instead, it attempts to discover patterns in the data. For example, assume that you have collected data describing the heights and weights of people. An example of an unsupervised learning problem is dividing the data points into groups. A program might produce groups that correspond to men and women, or children and adults [34].

(iii) Semi-supervised learning problem: It make use of both supervised and unsupervised data. For example, a reinforcement learning program that learns to play a side-scrolling video game such as Super Mario Bros. may receive a reward when it completes a level or exceeds a certain score, and a punishment when it loses a life. However, this supervised feedback is not associated with specific decisions to run, avoid Goombas, or pick up fire flowers[34].

The collection of examples that comprise supervised experience is called a training set. A collection of examples that is used to assess the performance of a program is called a test set. Two of the most common supervised machine earning tasks are classification and regression. In classification tasks the program must learn to predict discrete values for the response variables from one or more explanatory variables. That is, the program must predict the most probable category, class, or label for new observations. In regression problems the program must predict the value of a continuous response variable. A common unsupervised learning task is to discover groups of related observations, called clusters, within the training data. This task, called clustering [34].

By giving the training to the algorithm, different human activities can be classified and recognize. This task, called Classification. Fig.3 represent the approach based taxonomy for HAR [35].

ISSN: 2454-4248
462 – 471

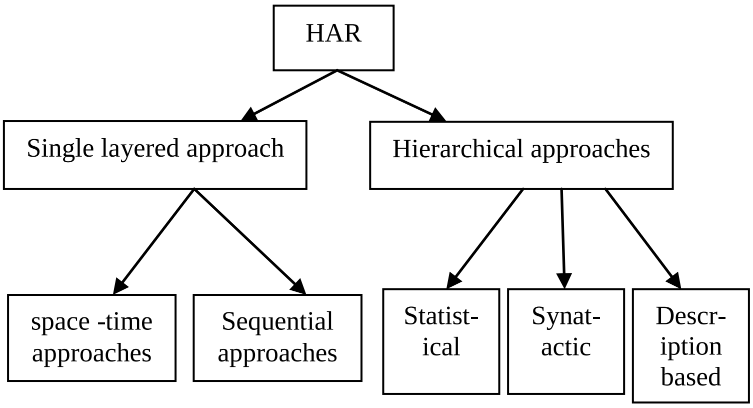

Fig. 3 Approach based taxonomy for HAR.

(A) Single-layered approaches

That represent and recognize human activities directly based on sequences of images. Due to their nature, single-layered approaches are suitable for the recognition of gestures and actions with sequential characteristics. Single-layered approaches are again classified into two types depending on how they model human activities: space-time approaches and sequential approaches. Space-time approaches model a human activity as a particular 3-D volume in a space-time dimension or a set of features extracted from the volume. The video volumes are constructed by concatenating image frames along a time axis, and are compared to measure their similarities [35].

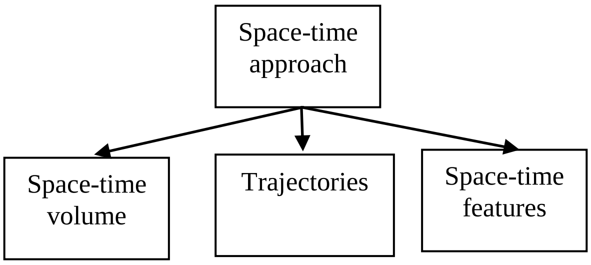

Fig. 4 Classification of space-time approach.

On the other hand, sequential approaches treat a human activity as a sequence of particular observations. Sequential approaches are the single-layered approaches that recognize human activities by analyzing sequences of features. They consider an input video as a sequence of observations (i.e. feature vectors), and deduce that an activity has occurred in the video if they are able to observe a particular sequence characterizing the activity. Sequential approaches convert a sequence of images into a sequence of feature vectors by extracting features (e.g. degrees of joint angles) describing the status of a person per image frame. Once feature vectors have been extracted, sequential approaches analyze the sequence to measure how likely the feature vectors are produced by the person performing the activity. If the likelihood between the sequence and the activity class (or the posterior probability of the sequence belonging to the activity class) is high enough, the system decides that the activity has occurred [35].

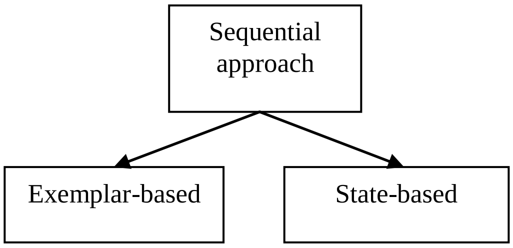

Fig. 5 Classification of Sequential approach.

(B) Hierarchical approaches:

The main idea of hierarchical approaches is to enable the recognition of high-level activities based on the recognition results of other simpler activities. The motivation is to let the simpler sub-activities which can be modeled relatively easily to be recognized first, and then to use them for the recognition
of higher-level activities. In general, common activity patterns of motion that appear frequently during high-level human activities are modeled as atomic-level (or primitive-level) actions, and high-level activities are represented and recognized by concatenating them hierarchically. The advantage is a result of two abilities of hierarchical approaches: the ability to cope with less training data, and the ability to incorporate prior knowledge into the representation. These approach classified into three category: Statical, Syntactic, and Description - based [35].

Statistical approaches use statistical state-based models to recognize activities. In the case of hierarchical statistical approaches, multiple layers of state-based models are used to recognize activities with sequential structures. A Syntactic approaches model human activities as a string of symbols, where each symbol corresponds to an atomic-level action. Similar to the case of hierarchical statistical approaches, syntactic approaches also require atomic-level actions to be recognized first, using any of the previous techniques.

A description-based approach is a recognition approach that explicitly maintains human activities' spatio-temporal structures. They represent a high-level human activity in terms of simpler activities composing the activity (i.e. sub-events), describing their temporal, spatial, and logical relationships. That is, description-based approaches model a human activity as an occurrence of its sub-event (which might be composed of their own sub-events) that satisfies certain relations [35].

There are many Classification methods used to identify and recognize Human activities. Some of the Classification methods and, it's advantages and disadvantages are presented in bellowing Table.3

ISSN: 2454-4248
462 – 471

TABLE II
SOME CLASSIFICATION METHOD FOR HUMAM ACTIVITY RECOGNITION

Classification Method	Description	Advantage	Disadvantage
HMM (Hidden Markov model)	A model used for modeling generative sequences by a set of observable Sequences. The main aim of this model is to determine the hidden state sequence from the observed output sequence.	Used to model complex activities.	Incapable of capturing transitive dependencies to its assumptions.
SVM (Support vector machine)	SVM is a supervised model that analyzes data and recognizes patterns. SVM classifies the data into a high dimensional space where a hyperplane is created for separation. On each side of this hyperplane, two separate hyperplanes are created. SVM tries to find the separating hyperplane which maximizes the distance between the two parallel hyperplanes.	Perform linear and non linear classifi-cation.	Higher computation burden for the constrained optimization program-ming used in the learning phase
NN (Neural network)	Neural Network consists of collection of inputs and processing units known as neurons. Neurons are arranged into three layers i.e. the input layer, the hidden layer and the output layer. The increasing number of hidden layer neurons results in increase of classification rate. For each activity, one neural network is trained.	Perform task that a linear program can't. High level of abstraction.	Time consuming.
ANN (Artificial Neural network)	Artificial neural networks are algorithms that can be used to perform nonlinear statistical modeling and provide a new alternative to logistic regression. ANN with Back propagation (BP) learning algorithm is widely used in solving various classification and forecasting problems. Even though BP convergence is slow but it is guranteed.	Requiring less formal statistical training, ability to implicitly detect complex nonlinear relationships between dependent and independent variables, ability to detect all possible interactions between predictor variables, and the availability of multiple training algorithms.	Disadvantages include its "black box" nature, greater computational burden, proneness to overfitting, and the empirical nature of model development.
DBN (Dynamic Bayesian network)	DBN is a generative model widely used for modeling temporal events in human activity recognition. The parameters of the DBN models are usually learned through maximizing likelihood or expected likelihood. It is an extension of a HMM, composed of multiple conditionally independent hidden nodes that generate observations at each time frame directly or indirectly.	provides a more general system model	inevitable need of very long training time, higher complexity and computation cost
K-NN (K - Nearest neighbor)	The K-nearest neighbor (K-NN) algorithm is a classification method based on the K, a predefined constant, closest training data in the feature space. A point/vector is classified to one label, which is the most frequent label among K nearest training points/vectors.	Simplest classification technique, Computation time for testing phase is independent of the number of classes, K-NN is robust in the search space even for nonlinearly separable data.	Classification performance is Sensit- ive to the selection of K.
Decision Tree	Classification model that breaks the dataset into smaller subsets while at same time an associated decision tree is incrementally developed	Learns very fast	Has trouble dealing with noise in data
DTW (Dynamic Time Warping)	The DTW a method for measuring similarity between two temporal sequences, which may vary in time or speed, is one of the most common temporal classification algorithms due to its simplicity; however, DTW is not appropriate for a large number of classes with many variations.	The advantage of DTW is that it is fast and easy.	needs extensive templates for various situations, resulting in high computation cost to match with these extensive templates.
Binary Tree	A binary tree is a tree structure with a maximum of two children for each internal node. For each node of the binary tree classifier, a Bayesian classifier is used and the likelihood functions are modeled and	Binary tree classification is simple and fast	The separation rules in each node are difficult to be general for other cases, making it difficult for complex scenarios.

	systematically learned as Gaussian mixtures.		
BoW (Bag of Words)	Probabilistic activity prediction model used to construct visual word histograms.	Easy to implement and use	Ignores spatial relationship among patches important for representation

Discussion about Classification methods

In above Table.2, There are some of the Classification methods are presented, which are used in different applications of Human activity Recognition to Recognize different level of Human activities. According to Ben companjon [36] When using HMM, stationary activities like sitting, reading, typing, lying down, Classification accuracy was much higher compared to using feature vectors. To effectively model the temporal causality of human activity, the HMMs with left-to-right state-transition structures are commonly used. However, many human actions exhibits the quasi-period cycles of body movements, which can't be easily modeled by basic HMMs. To overcome this problem, Thus et al.[37] proposed Cyclic HMM(cHMM), which is left-to-right HMM model with a return transition from the ending state to the beginning state. Because of inability of motions of two or more agents in basic HMM, Oliver et al.[38] construct a variant of the basic HMM, the Coupled HMM(cHMM), to model human-human interaction. Duong et al.[39] introduce the Switching Hidden semi-Markov model(S-HSMM) to recognize human daily life activities detect abnormality. Luo et al.[40] said that The sport behaviors are effectively recognized by using DBNs, which can generate a hierarchical description for video events, including bowing, downhill skiing, golf swing, pitching & ski jumps. Park and aggrwall [41] used DBN to recognize gestures such as ' stretching arm', 'turning a head left', by constructing a tree-structured DBN to take advantage of the dependent nature among body part's motion. The HMMs & DBNs suitable for modeling sequential relationship, not concurrent relationships. For example, HMMs & DBNs are difficulty modeling the relationship of an activity A occurred 'during', 'started with', or ' finished with' an activity B[35].

The SVM is one of the most popular margin-based supervised classifier in the pattern recognition.

Laptev et al. [42] use a nonlinear SVM with a multi-dimensional Gaussian kernel for recognition of various natural human activities successfully, including Answer Phone, GetoutCar, hand-shake, sit-down, sit-up & standup. Many classifier face the constraints of the long training time, and large size of the feature vector. K.G. Manosha et al.[43] proposed SVM classifier , on an existing spatio-temporal feature descriptor resolves these problems in HAR. To perform action classification of Weizmann Dataset, they train Multi-class SVM classifier with labeled action descriptor, They get 100% Recognition Rate and the system is consistently superior in regard to computationally time. Song et al. [44] use Neural networks to accurately detect falling , so that fall detection system carried at the left side of the waist can quickly and report a fall. A neural network classifier has been designed for human activity data, Frquency domain feature & fast training algorithm Levenberg marquardt algorithm was used for training. The designed neural network is giving improved mean classification rate compared to other results without the neural networks [45]. In pattern Recognition field, K-NN is one of the most important non-parameter algorithm and supervised algorithm accuracy of algo. can be severally degraded by the presence of noisy or irrelevant features.

K-NN is a type of instance-based learning, in which an object is classified by a majority of vote of its neighbors, with the object being assigned to the most common against its K nearest neighbors [46]. Jaideep chawla et al. [47] said that K-NN is a feasible choice as a classifier in limited dataset condition but the certain limitation need to be taken into consideration is data was collected mostly in a fitness studio with subjects who are well trained & perform the movements with the correct form. The Tree Classifier can flexibly implement different decision rules at its internal nodes, and can be adapted from a population based model. chien et al. proposed a system was tested using seven subjects, Each subject performed fourteen different activities typical of daily life. Using leave-one-out cross validation , decision tree produced average classification accuracies of 89.9% .

DTW is a similarity measure for two sequences, possibly with different length and different rate of occurrence. However , DTW issues of extensive templates for dataset, resulting in high computational cost. For example, speech recognition of the English character alphabets , it might need thousands of templates based on different accents. Darrell and pentland proposed a DTW based gesture recognize methodology using view models to represent the dynamic of articulated objects. Their system recognize 'good-bye', 'hello', gestures , and was able to distinguish them from other gestures such as a 'come closer' gestures. The bag-of-words approach were particularly successful for simple periodic actions.

VIII. Conclusion

In this Paper, We have present the detailed survey on each step as well as methods of Human activity Recognition. A Quality of data affect the methods used for action recognition. Therefore, The Quality of data must be an efficient , for this purpose we describe different pre-processing methods. Segmentation is important step, because in our consideration ,Recognition based on the related body parts involved in a particular action is more advantageous than considering the the entire body, for this purpose Segmentation of images or frames is used. Motion analysis is an key procedure for atomic human activity & social Role identification. The Feature extraction is possibly the most important part of the activity recognition problem since Classification can be handled by any existing machine learning algorithm if the features are robust. We have present different Feature extraction methods based on Time, Frequency, and Physical feature. In which, Time & Frequency features are more robust. HAR based on Single layer approach in which perform recognition directly from the raw data, is only efficient for low level activities like gestures and sign language recognition , where Hierarchical approach performs recognition using previously recognized atomic level activities, Therefore it is more suitable for High - level human activity Recognition. A detailed survey on Classification method shows that a Traditional classification method like HMM, SVM, DBN, K-NN,etc.. are less efficient compared to thier extended version like CHMM, S-HMM, Multi class SVM, Hierarchical DBN, etc.. so we concluded that we can get efficient accuracy result using the Fusion of Classification techniques rather than using only one classification technique for HAR.

.REFERENCES

[1] K.K. Roudposhti, J. Dias, P. Peixoto, "*A Multilevel Body Motion-based Human Activity Analysis Methodology*", IEEE Transaction on Cognitive and Development Systems, Vol 9, Issue 1, pp.16-29,2017.

[2] B. AY, M. Karakose , *"Motion Classification Approach Based on Biomechanical Analysis of Human Activities"*, IEEE Conference on Computational Intelligence and Computing Research, India, pp.1-8,2013.

[3] C. Hong, S. Xiao, Z. Tan, *"Real-Time Motion Recognition Based on Skeleton Animation"*,Image and Signal Processing (CISP), China, pp 1648-1652,2012.

[4] C. Zou, Z. Liu, *"Behavior Classification Mrthod based on Skeleton Model from Video images"*, International Conference on Computer Science and Information Technology, Singapore, pp.309-312, 2008 .

[5] S. Oprisescu, C. Burlacu, *"Action Recognition using Time of Flight Cameras"*, 8th International Conference on Communications (COMM),pp.153-156,2010.

[6] Wanare, Yogesh J., and P. R. Deshmukh. "A Survey on Human Action
Recognition using Scale Invariant Feature Transformation."

[7] Rautaray, Siddharth S., and Anupam Agrawal. "Vision based hand gesture recognition for human computer interaction: a survey." Artificial Intelligence Review 43.1 (2015): 1-54

[8] E.J. Amirbandi, G. shamsipour, *"Exploring Methods and Systems for Vision Based Human Activity Recognition"*, 1st Conference on Swarm Intelligence and Evolutionary Computation, Iran,2016.

[9] D.Barman, U.M. Sharma, *"A study on Human Activity Recognition from Video"*, 3rd International Conference on Computing for Sustainable Global Development (INDIACom), pp.2832-2835,2016.

[10] Vishwakarma, Sarvesh, and Anupam Agrawal. *"A survey on activity recognition and behavior understanding in video surveillance."* The
Visual Computer 29.10 (2013): 983-1009.

[11] Haritaoglu, I., Harwood, D., Davis, L.S.: W 4: *"Real-time surveillance of people and their activities"*. IEEE Trans. Pattern Anal. Mach. Intell.
22(8), 309–330 (2000)

[12] K. Kim, T.H. Chalidabhongse, D. Harwood, *"Real-time foreground-background segmentation using codebookmodel"*,Real time Imaging, 2014.

[13] Wren CR, Azarbayejani A, Darrell T, Pentland A. Pfinder: *"realtime
tracking of the human body"*. IEEE Transactions on Pattern Analysis and Machine Intelligence 1997;19(7):780–5.

[14] Horprasert T, Harwood D, Davis LS.*" A statistical approach for real-time robust background subtraction and shadow detection"*. IEEE Frame-Rate Applications Workshop, Kerkyra, Greece; 1999.

[15] Stauffer C, Grimson WEL. *"Adaptive background mixture models
for real-time tracking"*. IEEE International Conference on Computer Vision and Pattern Recognition 1999;2:246–52.

[16] Bayona, Álvaro, Juan C. SanMiguel, and José M. Martínez. "Stationary
foreground detection using background subtraction and temporal difference in video surveillance". Image Processing (ICIP), 2010 17th
IEEE International Conference on. IEEE, 2010.

[17] Shibata, M., Yasuda, Y., Ito, M.: *"Moving object detection for active
camera based on optical flow distortion"*. In: Proc. of the 17th World Congress the International Federation of Automatic Control,
Seoul, Korea, pp. 14,720–14,725 (2008)

[18] V. Tiwari, D. Chaudhary, *"Foreground Segmentation using GMM Combined Temporal Differencing"*, International Conference on Computer, Communications and Electronics ,Jaipur,2017 .

[19] A. Chase, B. Sprecher, J. Hoffmire, *"Motion Detection and Segmentation Using Optical Flow"*.

[20] J. Suto, S. Oniga, P.P Sitar, *"Feature Analysis to Human Activity Recognition"*, International Journal of Computers, Communications & Control (IJCCC), 2017.

[21] N. Yala, B. Fergani, A. Fleury, *" Feature extraction for human activity recognition on streaming data"*, IEEE, France, 2015

[22] M.A. Hall, L.A. Smith, *"Feature Selection for Machine Learning: Comparing a Correlation-based Filter Approach to the Wrapper"*, Twelfth International FLAIRS Conference, New Zeland,1999.

[23] R. Damasevicius, M. Vasiljevas,*"Human Activity Recognition in AAL Environments Using Random Projections"*, Research Article, Poland,2016.

[24] Z.M. Hira, D.F. Gillies, *"Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data"*, Review Article, UK,2015.

[25] R. Diaz-Uriarte and S. Alvarez de Andres, *"Gene selection and classification of microarray data using random forest,"* BMC Bioinformatics,vol.7,article3,2006.

[26] Uddin, M. Z., Lee, J. J., Kim, T.-S., *"Independent shape component-based human activity recognition via Hidden Markov Mode"*l, Journal of Applied Intelligence (Springer). DOI: 10.1007/s10489-008-0159-2.,Netherlands,2009

[27] http://en.wikipedia.org/wiki/Independent_component_analysis

[28] Kwak K-C, Pedrycz W, *"Face recognition using an enhanced independent component analysis approach"*,IEEE Trans Neural Netw 18(2):530–541, 2007.

[29] C. Ding and H. Peng, *"Minimum redundancy feature selection from icroarray gene expression data,"* in Proceedings of the IEEE Bioinformatics Conference (CSB '03), pp. 523–528, IEEE Computer Society, Washington, DC, USA, August 2003.

[30] M. A. Hall, *"Correlation-based feature selection for discrete and numeric class machine learning,"* in Proceedings of the 17th International Conference on Machine Learning (ICML '00),pp.359–366, Morgan Kaufmann, San Francisco, Calif, USA, 2000.

[31] Y. Wang, I.V.Tetko, M.A. Halleta., *"Geneselectionfrom microarray data for cancer classification—a machine learning approach,"* Computational Biology and Chemistry,vol.29,no.1,pp.37–46,2005

[32] M. R.Sikonja and I. Kononenko*, "Theoretical and empirical analysis of Relief F and RRelief F,"* Machine Learning,vol.53,no. 1-2, pp. 23–69, 2003.

[33] Gavin Hackeling, "*Mastering Machine Learning with scikit-learn",* Packt Publishing Ltd,UK,2014, 978-1-78398-836-5.

[34] J.K. Aggrwal, M.S.Ryoo, "*Human Activity Analysis: A Review",* ACM Journal.

[35] B. Companjen, *"Classification methods for activity recognition",* 11thTwente Student Conference on IT, Enschede 29th June, 2009

[36] Hoang, L.U.T.; Ke, S.; Hwang, J; Tuan, P.V.; Chau, T.N. *"Quasi-periodic action Recognition from Monocular Videos via 3D Human Models and Cyclic HMMs"*. In Proceedings of IEEE International Conference on Advanced Technologies for Communications (ATC), Hanoi, Vietnam, 10–12 October 2012; pp. 110–113.

[37] Oliver, N. M., Rosario, B., and Pentland, A. P. 2000. *"A Bayesian computer vision system for modeling human interactions"*, IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 8, 831-843.

[38] Duong, T.V.; Bui, H.H.; Phung, D.Q.; "Venkatesh, S. *Activity Recognition and Abnormality Detection with the Switching Hidden Semi-Markov Model".* In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), San Diego,CA, USA, 20–25 June 2005; Volume 1, pp. 838–845.

[39] Luo, Y.; Wu, T.; Hwang, J. *"Object-based analysis and interpretation of human motion in sports video sequences by dynamic Bayesian networks". Comput Vis. Image Underst.* 2003, *92*,196–216.

[40] Park, S. and Aggarwal, J. K. 2004. *"A hierarchical Bayesian network for event recognition of human actions and interactions. Multimedia Systems* ",10, 2, 164-179.

[41] Laptev , J. Goulermas, L. Kenney, and D. Howard, *"Activity identification using body-mounted sensors a review of classification techniques*". Physiological measurement, (30), Apr 2009.

[42] *K..G.Manosha, R..Rodrigo, "*Faster Human Activity Recognition with SVM*"*, The International Conference on Advances in ICT for Emerging Regions – ICTer 2012 : 197-203

[43] S. K. Song, J. Jang, and S. Park, *"A phone for human activity recognition using triaxial acceleration sensor*". In Digest of Technical Papers - IEEE International Conference on Consumer Electronics, 2008.

[44] A. Sharma, Y. Lee, W.chung, *"High Accuracy Human Activity Monitoring using Neural network ",* Third 2008 International Conference on Convergence and Hybrid Information Technology,Korea, 2008.

[45] S. Kaghyan, H. Sarukhayan, "*Activity recognition using K-nearest neighbor algorithm on smartphone with tri-axial accelerometer",* International Journal "Information Models and Analyses" Vol.1 / 2012.

[46] J.Chawla, M.Wanger, *"Using Machine Learning Techniques for User Specific Activity Recognition"*, Proceedings of the Eleventh International Network Conference,Germany,2016.

[47] Wren, C.R.; Azarbayejani, A.; Darrell, T.; Pentland, A.P. Pfinder, *"Real-time tracking of the human body*". *IEEE Trans. Pattern Anal. Mach. Intell.* 1997, *19*, 780–785..