

Person Identification using Skeleton Information from Kinect

Aniruddha Sinha
 Innovation Labs,
 Tata Consultancy Services
 Kolkata, India
 aniruddha.s@tcs.com

Kingshuk Chakravarty
 Innovation Labs,
 Tata Consultancy Services
 Kolkata, India,
 kingshuk.chakravarty@tcs.com

Brojeshwar Bhowmick
 Innovation Labs,
 Tata Consultancy Services
 Kolkata, India
 b.bhowmick@tcs.com

Abstract—In recent past the need for ubiquitous people identification has increased with the proliferation of human-robot interaction systems. In this paper we propose a methodology of recognizing persons from skeleton data using Kinect. First a half gait cycle is detected automatically and then features are calculated on every gait cycle. As part of new features, proposed in this paper, two are related to area of upper and lower body parts and twelve related to the distances between the upper body centroid and the centroids derived from different joints of upper limbs and lower limbs. Feature selection and classification is performed with connectionist system using Adaptive Neural Network (ANN). The recognition accuracy of the individual people using the proposed method is compared with the earlier methods proposed by Arian et. al and Pries et. al. Experimental results indicate that the proposed approach of simultaneous feature selection and classification is having better recognition accuracy compared to the earlier reported ones.

Keywords—*Person identification; gait recognition; adaptive artificial neural network(ANN); Kinect; connectionist system*

I. INTRODUCTION

Automatic person identification is one of the important factors in Human-robot interaction based applications [1]. Successful identification of individual gait pattern can certainly help a machine to take and control different actions on the basis of previous interaction with that same individual. The interaction by the robot can be personalized upon identification of an individual [2][3]. Biometric identification, using different behavioral patterns like gait, keyboard typing, lip movement etc. or physiological signatures like voice/speech, face, iris, fingerprint etc. [4], is one of the common means of automatic person recognition. Over the last few decades, extensive research work has already been done on different biometric modalities like face, iris scan and fingerprint recognition [5]. However, other biometric characteristics like gait, skeleton data, EEG are relatively less considered [8]. Moreover, it is very difficult to get iris, fingerprint, face or audio (voice/speech) related biometric information (at the recognizable resolution) from a large distance and without user's direct co-operation. Human gait recognition has great advantages in recognition from low resolution images where other biometric techniques are not suitable because of insufficient pixel information [6]. Psychophysical studies [7] indicate that human being is capable of recognizing an individual reliably,

from his/her style of walking or movement. Another advantage of gait pattern identification is that it is very much hard to hide or imitate.

For gait information processing, we need video sequence of walking person, where at least one complete gait cycle is present. A gait cycle is starting with one foot forward and ending with same foot forward. Each and every gait has two main components [9] namely, a structural component or physical build of a person (e.g. height, length of limbs etc.) used to derive static features and motion dynamics during gait cycle used to derive dynamic features.

In this paper, we use Microsoft Kinect sensor to derive the skeleton data for recognizing of individuals. The main contribution of this paper is given below:

- Automatic gait cycle detection algorithm from skeleton data of Kinect as opposed to manual gait cycle detection [22].
- Proposal of new static and dynamic features related to area and distance.
- Methodology to use ANN based connectionist framework for feature selection and supervised learning algorithm [30] to determine the important features.

The proposed area features include the area spanned by the lower and upper parts of the body as physical/structural features. The distance features include relative distances between body-centroid and the centroids formed by the joints of upper and lower limbs. Comparison of recognition accuracy is done between the proposed approach and the one mentioned in [21] and [22].

This paper is structured as follows: Related work is presented in Section II. Proposed method for Kinect data recording setup and feature extraction are given in Section III. The supervised learning and ANN based feature selection is given in Section IV. The results are given in Section V followed by conclusion and acknowledgement.

II. Related work

Existing gait recognition methods can be categorized under two categories namely model based and model free approaches [10]. In the model based approach, gait signature is derived from modeling and tracking different parts of the body (like legs, limbs, arms etc) over time and then it is used for person recognition and identification. Primary model based approaches consider only static parameter for recognition [11]. BenAbdelkader et al. [12] used structural

parameter like stride and cadence, of walk-cycle for identification. Yam et. al [13] built a structure and motion based model of legs to indiscriminate walking and running gait pattern using biomechanics of human and pendulum motion. Though this approach is view-invariant and scale independent, but it is computationally expensive and very much sensitive to quality of gait sequences [10]. Model free approaches mainly characterize gait pattern by observing how shape of the silhouette of an individual changes over time or by considering entire motion dynamics of the person. Sarkar et. al [14] propose a baseline algorithm using series of gait silhouettes as feature. Gait Energy Image (GEI) [15] and Motion Energy Image (MEI) [16] are also used for individual identification. Though GEI is robust but it lacks dynamic gait information. Liu and Zheng [17] propose an algorithm which combines both spatial and temporal information. To solve the problem of gait incompleteness, Chen et. al [18] suggests FDEI algorithm. Xue et al. [19] use wavelet decomposition of GEI for infrared gait recognition. In many of these approaches, width of the outer silhouette contour has been considered as a good feature candidate. Model free approaches are relative easy but not robust with view point and scale. For direct gait classification, many researchers use KNN, SVM and TSVM. Dynamic Time Wrapping (DTW) is a very popular technique for gait pattern matching. Sundaresan et al. [23] successfully develop a HMM based framework for gait recognition. Recently [21] has employed 3D skeleton based approach for recognizing individual person from their walking gait pattern. They have used Microsoft Kinect to capture skeleton data and used K-means as an unsupervised learning algorithm for clustering. Pries et. al. [22] also does a comparable job of skeleton based person recognition using Kinect. They have used 11 static and 2 dynamic features for classification and compared 1R, C 4.5 decision tree and Naive Bayes classification results. To best of our knowledge there have been no work done on the feature selection of Kinect data for person identification. PCA [20] and LDA [24] are most commonly used feature selection tool in this field. More recently, Mu and Tao [25] use DLA to reduce biologically interdependent features. In this paper we investigate the importance of various features using the ANN based feature selection method.

III. PROPOSED METHOD

In this paper we propose a skeleton based approach (provided by Microsoft Kinect Sensor) for gait recognition and person identification. Kinect provides real time human skeleton information of 20 skeleton joints. We have used Microsoft Kinect SDK for recording the data for human skeleton at 30 frames per second (fps).

Our system is implemented mainly in three steps - recording of skeleton data for side walking pattern using Kinect (section III.A), feature generation over half gait cycle (section III.B) and supervised learning and feature selection to identify an individual (section III.C).

A. Skeleton Data Recording

We recorded the 20 joints of skeleton data for a person, keeping position of Kinect fixed throughout our experiment. In our experiment, side way walking pattern of an individual on an arbitrary path is considered. In each recording session, we have taken multiple side-walks (from left to right and right to left), where the distance of the persons from the Kinect ranged between 6 feet to 10 feet (Figure 1). This enables us to test the identification accuracy in a scenario where the distance of the subject changes over time.

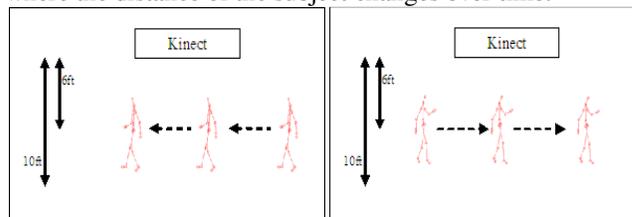


Figure 1. Recording setup using Kinect -- (a) showing right to left walk and (b) showing left to right walk

B. Feature Generation

Salient feature generation from skeleton data is an important step to discriminate individual characteristics. In case of walking, half gait cycle provides all the meaningful information required for identification. Thus initially, the half gait cycles need to be identified from the skeleton data and then generate the features per half-gait cycle.

Computation of Half-Gait Cycle

The side walk movement pattern of a person is shown in Figure 1. We have considered x-axis as horizontal axis, parallel to ground. Figure 2 shows the changes in left and right ankle position over time (frame number) indicating an individual is moving multiple times from left to right and right to left in front of the Kinect.

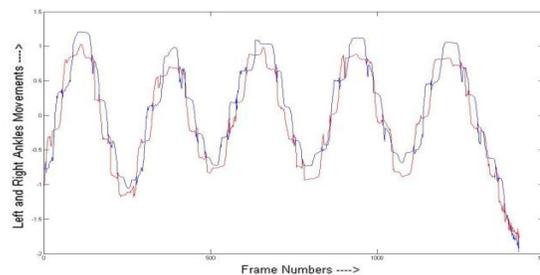


Figure 2. Left (marked in Red color) and Right (marked in Blue color) Ankle movement profile

Horizontal distance i.e., difference between x-co-ordinates of left and right leg ankles is computed for each recorded frame using (1).

$$distance_k = |x_{leftankle} - x_{rightankle}| \quad (1)$$

For all $k = [1, N]$, where $N =$ total number of frames for individual sidewalk ($N > 1$).

This distance vector is smoothed using moving average algorithm with small window size. We have detected the transition from negative slope to positive slope to identify all

the local minima. The half gait cycle is considered as the frames between two consecutive local minima. Figure 3 shows the square of $distance_k$ for several frames, in dotted green color and the half gait cycle boundaries are shown with vertical partitions in red color.

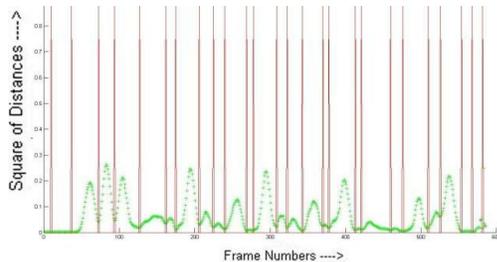


Figure 3. Detection of Half Gait Cycle

Feature Vector

The features that can be generated from the skeleton data of 20 joints can broadly be classified as static and dynamic features. The static features are based on the distances between the adjacent physical joints as shown in the following equation

$$d_{ij} = \sqrt{\|x_i - x_j\|^2 + \|y_i - y_j\|^2 + \|z_i - z_j\|^2}$$

where the joints i and j are directly connected and adjacent to each other. The dynamic features include the variation due to the joint movements and their orientations with respect to the other joints which are not adjacent to each other.

In [21] Adrian et. al have proposed eighteen features related to the angle changes in different joints below hip. These are mainly dynamic features where they have taken the mean, max and standard deviation of the three angles for each leg. We have considered these features for comparing the accuracy of individual identification. In [22] Preis et. al have proposed 12 static features (distances between joints) and 2 dynamic features (stride length and speed) for people identification. These are also considered for comparison.

In this paper we have proposed a two area related features and a set of hybrid features which are a mixture of static and dynamic information. Out of these, the first area feature f_{au} is static information of an individual and changes insignificantly while walking or movement. The second area feature f_{al} is a dynamic feature and changes significantly while walking. The proposed twelve hybrid features are distance features f_D which are related to static and dynamic information of an individual.

Area Features: Area occupied by upper and lower part of the body during side walk is the key distinguishing factor as the spread of hand and leg for each person is different from others. So we have considered area of upper and lower body during side walk as one of our features.

A reasonable way to compute area of upper or lower part of a body is to select N joints ($3 \leq N \leq 20$), so that N joints form a closed polygon. If co-ordinate of $i^{th} (\forall i \in N)$ joint is (x_i, y_i) , then the area enclosed by N joints, as projected on Z -plane, is

$$A = \frac{1}{2} \sum_{i=0}^N (x_i * y_{i+1} - x_{i+1} * y_i) \quad (2)$$

Using (2) we have calculated area of upper body (f_{au}) and lower body (f_{al}). The joint considered for upper body area are shoulder centre, shoulder left, hip left, hipcentre, hip right and shoulder right. The joints considered for lower body area are hip centre, hip right, knee right, ankle right, ankle left, knee left and hip left. The final area feature vector is the mean area values in a half-gait cycle, given as $f_A = \{f_{au}, f_{al}\}$ is in R^2 space.

Distance Features: The change in distances of skeleton joints with respect to the centroid of the upper body is unique for a person and is easily recognizable by a human brain. Thus, the Euclidean distances between centroid of different parts of the body with respect to the centroid of the upper body are indeed a set of good candidate as features. Here, we have considered only four distances between upper body centroid and to the centroids of both hands and legs, separately.

For a closed polygon of N vertices, the centroid \vec{X}_c can be computed as

$$\vec{X}_c = \frac{1}{N} \sum_{i=0}^N N(x_i, y_i, z_i) \quad (3)$$

Using (3), we have calculated the centroid for the following:

- upper body (enclosed by shoulder centre, shoulder left, hip left, hip right and shoulder right)
- right hand (enclosed by shoulder right, elbow right and wrist right)
- left hand (enclosed by shoulder left, elbow left and wrist left)
- right leg (enclosed by hip right, knee right and ankle right)
- left leg (enclosed by hip left, knee left and ankle left)

The Euclidean distances f_{ci} between upper body centroid (x_c, y_c, z_c) and other centroids (x_i, y_i, z_i) can be written as given in (4) where, $i=1$ for left hand, $i=2$ for right hand, $i=3$ for left leg and $i=4$ for right leg.

$$f_{ci} = \sqrt{\|x_c - x_i\|^2 + \|y_c - y_i\|^2 + \|z_c - z_i\|^2} \quad (4)$$

$\forall i \in \{1,2,3,4\}$

Thus the final hybrid distance feature vector is in R^{12} space given by (5)

$$f_D = \{f_{ci}^{mean}, f_{ci}^{stddev}, f_{ci}^{max}\} \quad (5)$$

where, $f_{ci}^{mean} = \text{mean}(f_{cik}), f_{ci}^{stddev} = \text{stddev}(f_{cik}), f_{ci}^{max} = \text{max}(f_{cik}), \forall k \in \text{halfgait}$

Figure 4 shows Euclidean distance between upper body centroid and right hand centroid.

We have also considered all the features ($f_{[21]}$) and ($f_{[22]}$) reported in the papers [21] and [22] respectively however, instead of taking angle between foot and ground, angle between ankle and ground is used because left and right foot co-ordinates are relatively more inclined to noise. Thus finally the combined feature vector is in R^{46} space as given by (6).

$$f = \{f_{[21]}, f_A, f_D, f_{[22]}\} \quad (6)$$

The feature data are the normalized to zero mean and unit co-variance across all the features before using for training and recognition.

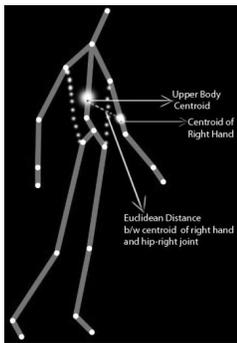


Figure 4. Euclidean distance between upper body centroid and right hand centroid

IV. SUPERVISED LEARNING

We have used Adaptive Neural Network (ANN) and Naive Bayes classifier for supervised learning. Pries et. al. in [22] have found that Naive Bayes performs better compared to 1R and C4.5 decision tree. Hence we have excluded them in our experiment.

A. Naive Bayes

The Naive Bayes classifier is based on Bayes' law. It is a probabilistic classifier which assumes the statistical independence of the features. This assumption is violated in the case of features extracted from the human gait cycle as clearly indicated by [22], however, it performs the best as reported with the dataset tested in [22]. Thus we also have considered the Naive Bayes classifier for comparing the proposed features along with one more non-linear classifier.

B. Multi layer Perceptron Models

Multi-layer Perceptron Model (MLP) is a well know ANN architecture [26]. This consists of an input layer, an output layer and one or more hidden layers. The excitation function of the neurons (nodes) used for the experiment is a uni-polar sigmoid function. The strengths of the connections between the nodes of different layers are termed as weights. These weights are initialized randomly and the final weights are obtained after running the back-propagation (BP) algorithm [27] iteratively over the training feature vectors. A sample structure of the ANN is shown in Figure 5 where a single hidden layer is shown but in general multiple such hidden layers can be present. The weight matrix $\overrightarrow{W^1}$ corresponds to the weights between the hidden layer and the input layer. Similarly, $\overrightarrow{W^2}$ corresponds to the weight matrix between the hidden layer output layer and the output layer. The output of the neural network is denoted by the hypothesis function $h_w(f)$.

In our experiments we have considered only a single hidden layer. It is observed that increasing more number of nodes or increasing number of hidden layer is not improving the results and it was taking more time to execute. By

experimentation and using the method proposed by Hagiwara [29], we have chosen a single hidden layer with number of nodes as 25.

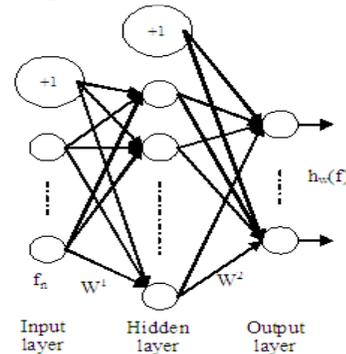


Figure 5. A sample structure of the ANN

The cost function of the neural network is given by

$$J(w) = \frac{1}{m} \sum_{i=1}^m \sum_{k=1}^K [-y_k^{(i)} \log((h_w(f^{(i)}))_k) - (1 - y_k^{(i)}) \log(1 - h_w(f^{(i)}))_k] + \frac{\lambda}{2m} [\sum_{j=1}^H \sum_{k=1}^N (W_{jk}^1)^2 + \sum_{j=1}^K \sum_{k=1}^H (W_{jk}^2)^2] \quad (7)$$

where various terms are described below:

- m is the number of training samples
- The input training features are $f_n^{(i)}$, where n = 1, 2 ... N and i = 1, 2, ... m. Here N is the number of features and m is the number of training samples. Thus is $f_n^{(i)} \in \mathbb{R}^N$ an N dimensional feature vector.
- K is the number of classes which is the number of nodes in the output layer. The class corresponding to the feature vector is a vector $y_k^{(i)}$ where k = 1, 2, ... K. For the class vector corresponding to pth class is (0, 0, .. 1, .. 0) where the pth entry is 1 and remaining are 0.
- H is the number of hidden layer.
- λ is the regularization parameter
- $\overrightarrow{W^1}$ and $\overrightarrow{W^2}$ are the weight matrices.

The steps for training the neural network is given in [30].

We have used K-fold cross validation to choose regularization parameter λ . Figure 6 shows the validation error with respect to λ .

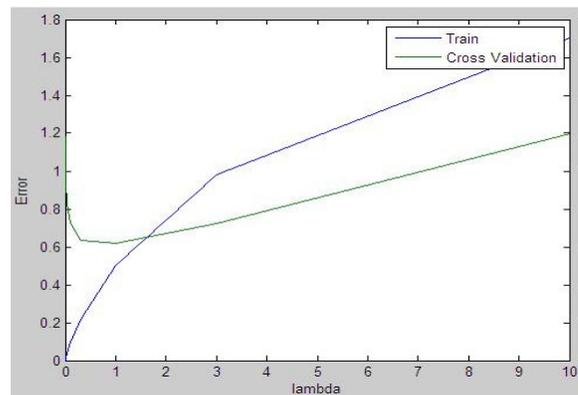


Figure 6. Validation curve of the ANN

The regularization parameter λ is selected as 0.6 which gives the minimum cross-validation error. Learning curve is used to verify that the training samples are enough to train the designed ANN structure (Figure 5).

C. Feature Selection using ANN

In this paper, we provide a methodology to use the connectionist system for feature selection as proposed by Pal et. al [30] in order to get more insight to the previous and new features. In [30], an "on-line" methodology is provided for simultaneous feature selection and classification using MLP. We have used the same approach; however the learning is done in "Batch" mode rather than "on-line" mode. For each iteration, in the "Batch" mode, all the training samples are fed to the neural network and the error in each layer are accumulated to compute the gradient values of the cost function with respect to the weights. Whereas in "on-line" mode, for each iteration, a single training sample is used to compute the error in each layer to compute the gradient values. After every iteration, the weights are updated using conjugate gradient descent algorithm to minimize the cost function. The ANN structure is shown in Figure 7 where a "Selection Layer" is added before the standard input layer.

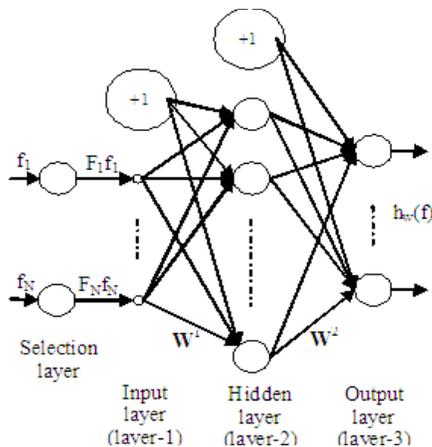


Figure 7. ANN Structure with Feature Selection

According to [30], the attenuation function in the selection layer is F and the argument for the attenuation function is M_n for the n^{th} feature. This attenuation function F is applied to the input features before passing to the normal neural network. During the learning process the arguments M_n gets updated such that the favourable features having higher discriminative properties get higher values in M_n while others get lower values. Thus the favorable features, termed as good features, get passed with higher strength while the others get lower strength. This process is called the selection of features which happens simultaneously with the training of the ANN.

In our experiment the attenuation function is chosen as a uni-polar sigmoid function given by (8).

$$F(M) = F_n = \frac{1}{1 + e^{-M_n}} \quad (8)$$

Thus the modified input to the input layer is $F_n f_n$ for $n = 1, 2 \dots N$, where N is the number of input features. Thus the

modified cost function for the batch mode is given by (9), where, $\vec{f}_a^i = \vec{f}^i * \vec{F}(\vec{M})$ is the i^{th} attenuated feature vector for the i^{th} training sample.

$$J(W, M) = \frac{1}{m} \sum_{i=1}^m \sum_{k=1}^K [-y_k^{(i)} \log((h_w(f_a^{(i)}))_k) - (1 - y_k^{(i)}) \log(1 - h_w(f_a^{(i)}))_k] + \frac{\lambda}{2m} [\sum_{j=1}^H \sum_{k=1}^N (W_{jk}^1)^2 + \sum_{j=1}^K \sum_{k=1}^H (W_{jk}^2)^2] \quad (9)$$

The gradients of the cost function (as in (9)) with respect to the weights ($\frac{\partial J}{\partial W^1}$ and $\frac{\partial J}{\partial W^2}$) and with respect to the arguments for the attenuation function ($\frac{\partial J}{\partial M}$) are computed using BP algorithm. In every iteration, all the training samples are used to accumulate the gradient values. These gradients are then used to calculate the updated weights and arguments for the attenuation function.

The derivation for updating \vec{W}^1 , \vec{W}^2 and \vec{M} are given in [30] which are based on on-line mode. For the batch mode, the gradient vectors in each layer are accumulated to compute the gradient of the cost function. The steps for updating the \vec{W}^1 , \vec{W}^2 and \vec{M} are given below:

1. Compute the feed-forward for the feature vector of the i^{th} training sample $\vec{f}^{(i)}$.
2. Calculate the error δ_k^3 in for each output node k in layer-3 by setting $\delta_k^3 = h(f^{(i)})_k - y_k^i$, where $y_k^i = 1$ if the training example belongs to k^{th} class, else $y_k^i = 0$.
3. For the hidden layer, calculate the error by setting $\delta_j^2 = O_j^2 \sum_i \delta_i^3 W_{ij}^2$, where O_j^2 is output of the excitation function of j^{th} node in the 2^{nd} layer and O_j^2 is the derivative of the same.
4. For the selection layer, calculate the error vector by setting $\delta_j^s = F_j' \sum_i \delta_i^2 W_{ij}^1$, where F_j' is the gradient of the attenuation function in the j^{th} input node.
5. The above W_{ij}^N are the weights between the i^{th} node of the $(N + 1)^{th}$ layer and the j^{th} node of the N^{th} layer.
6. For all the m training samples, accumulate the gradient vectors using

$$\Delta_{ij}^1 = \Delta_{ij}^1 + \delta_i^2 O_j^1 \quad \Delta_{ij}^2 = \Delta_{ij}^2 + \delta_i^3 O_j^2$$

$$\Delta_{jj}^s = \Delta_{jj}^s + \delta_i^s f_j$$

7. The gradient of the neural network cost function is obtained by the following equations,

$$\frac{\partial J(W, M)}{\partial W_{ij}^l} = D_{ij}^l = \frac{1}{m} * \Delta_{ij}^l, \text{ for } j = 0$$

$$\frac{\partial J(W, M)}{\partial W_{ij}^l} = D_{ij}^l = \frac{1}{m} * \Delta_{ij}^l + \frac{\lambda}{m} * W_{ij}^l, \text{ for } j \geq 1$$

$$\frac{\partial J(W, M)}{\partial W_{ij}^l} = D_j^s = \frac{1}{m} * \Delta_j^s$$

The λ is the regularization parameter.

8. Update the \vec{W}^1 , \vec{W}^2 and \vec{M} using the conjugate gradient descent algorithm [28].
9. After every update, the above steps 1 to 8 are repeated till the target number of iterations are over or the cost function has reached a local minima.

The correctness of the feature selection method in batch mode is verified by using the IRIS dataset [32]. This dataset has four features namely, sepal length (f_1), sepal width (f_2), petal length (f_3), and petal width (f_4) of Iris flower. It is observed that the features f_3 and f_4 get higher importance compared to features f_1 and f_2 . This is same as reported in [30] and [31]. Hence the design for the feature selection in batch mode for ANN is confirmed to be consistent to previously reported results. This ANN structure is used in our experiment to analyze the proposed features and previously reported features for human identification from Kinect based skeleton data.

V. EXPERIMENTAL RESULTS

The accuracy of recognizing an individual based on skeleton data is evaluated for the proposed features and compared with the features and results reported in earlier works [21], [22]. We have performed the experiments in four stages. Section V.A compares the proposed features with the one given in [21] for 5 subjects with manual gait data, section V.B gives the results for automatic gait detection, section V.C gives the effect of increase in subjects to 10 and comparison with [22] and section V.D gives the results for feature selection using ANN.

The results of performance comparison are given in terms of F-score, which is the harmonic mean of precision and recall and calculated by (10).

$$F_{score} = \frac{2 * precision * recall}{(precision + recall)} \quad (10)$$

A. Comparison with [21]

Initially, we have captured data for 5 individuals and compared the results with [21] as Adrian et. al have used 4 subjects in their experiment. The total number of half-gait cycles for 5 individuals is approximately 700 with an average of 140 half-gait cycles per individual.

As reported in [21], we perform k-means clustering on the features with k as 5 corresponding to 5 subjects. TABLE I shows the confusion matrix in percentage for the feature points as grouped in different clusters. It can be noted that the overall average accuracy of recognition is 25.2% as against the by chance accuracy of 20%.

TABLE I. CONFUSION MATRIX FOR 5 CLUSTERS

Person	Level A	Level B	Level C	Level D	Level E
1	19	40	10	31	0
2	19	20	25	31	5
3	23	20	20	34	3
4	23	21	21	32	3
5	24	25	16	31	4

In this experiment the half-gait cycles are manually extracted to keep the experimental environment similar to [21]. The results of the k-means clustering are compared with the ANN classification. Out of the 700 features vectors, 372 vectors are used for training and the remaining are used for testing. In our implementation, we have used ANN with one hidden layer of 25 nodes as shown in Figure 5.

A summary of the results for 5 subjects with manual half-gait cycle extraction is shown in TABLE II. This clearly indicates that if we take all the 32 features (14 new features ($f = \{f_A, f_D\}$) and 18 earlier features ($f_{[21]}$) proposed in [21]), the recognition accuracy is maximum at 86% which is substantially higher compared to by chance accuracy of 20%.

TABLE II. PERFORMANCE (F-SCORE IN %) COMPARISON OF DIFFERENT FEATURES AND CLASSIFIERS FOR 5 SUBJECTS WITH MANUAL GAIT CYCLE DETECTION

K-Means with $f_{[21]}$	ANN with $f_{[21]}$	ANN with $f = \{f_A, f_D\}$	ANN with $f = \{f_{[21]}, f_A, f_D\}$
25.2	34	69	86

B. Automatic Half-Gait detection

The accuracy of the half-gait detection algorithm (section III) is measured by comparing the ground truth generated by the manual gait detection. The precision of the algorithm is 1, recall is 0.79 and the F-score is found to be 0.88. The accuracy of the recognition of an individual for 5 subjects using automatic gait cycle detection is given in TABLE III. It can be seen that the performance has slightly degraded compared to the manual gait detection (TABLE II).

TABLE III. PERFORMANCE (F-SCORE IN %) COMPARISON OF DIFFERENT FEATURES AND CLASSIFIERS FOR 5 SUBJECTS WITH AUTOMATIC GAIT CYCLE DETECTION

ANN with $f_{[21]}$	ANN with $f = \{f_A, f_D\}$	ANN with $f = \{f_{[21]}, f_A, f_D\}$
33	68	64

A sample confusion matrix for 5 subjects is given in Figure 8 where half-gait cycle is detected automatically and only 14 new features $f = \{f_A, f_D\}$ are used.

```

fscoreTest = 0.682067
a   b   c   d   e  <-- classified as
40  11  36  0   1  | a
1   29  8   0   1  | b
33  9   20  0   3  | c
1   0   0   57  0  | d
0   1   1   0   76 | e
    
```

Figure 8. Confusion matrix for 5 persons using 14 new features

The results in the remaining section of this paper are based on automatic gait detection.

C. Effect on Increase in Subjects

It is important to analyze the effect in accuracy results as we increase the number of subjects. Pries et. al in [22] have used 9 subjects for their experiments where they have found Naive Bayes classifier to perform best on the 14 features, $f_{[22]}$ proposed by them. As the dataset used by [22] is not publicly available, we have generated a larger dataset of 10 subjects with 8 men and 2 women. The comparison of different features and classifiers are shown in TABLE IV. It can be seen $f = \{f_{[21]}, f_A, f_D\}$ with ANN classifier performs best having recognition accuracy of 55% which is much better than chance accuracy of 10%.

TABLE IV. PERFORMANCE (F-SCORE IN %) COMPARISON OF DIFFERENT FEATURES AND CLASSIFIERS FOR 10 SUBJECTS

ANN with $f_{[21]}$	ANN with $f = \{f_A, f_D\}$	ANN with $f = \{f_{[21]}, f_A, f_D\}$	Naïve -Bayes with $f_{[22]}$
27	52	55	51.3

A sample confusion matrix on the test data for ANN with $f = \{f_{[21]}, f_A, f_D\}$ is shown in Figure 9. It can be seen that the recognition for the subject "a" is worst affected whereas others are quite reasonable as seen from high values in the diagonal entries.

```

precision = 0.548295, recall = 0.552487, fscoreTest = 0.550383
confusion_mat
 a  b  c  d  e  f  g  h  i  j  <-- classified as
20  8  47  0  5  3  11  0  1  5  | a
 4  49 10  0  8  8  0  0  0  21 | b
 7  11 52  0  2  5  13  0  2  8  | c
 2  0  0  60  0  0  2  10 26  0  | d
 0  0  1  0  92  3  0  0  0  4  | e
10 10  5  0  2  44  2  0  0  27 | f
26  3  20  0  1  3  45  0  1  1  | g
 0  0  0  15  0  0  0  82  3  0  | h
 6  0  1  4  0  0  16  0  73  0  | i
 1  22  0  0  13 25  1  0  0  38 | j
    
```

Figure 9. Confusion matrix for 10 persons using 32 features

It is to be noted that the recognition reported here is based on every half-gait cycle, however in actual scenario the recognition will be done on a recording of a session where the decision will be taken on multiple gaits. Considering this fact, we can clearly say that the recognition is 90% (9 identified correctly out of 10) where the entry corresponding to the maximum number in the confusion matrix is taken to identify an individual.

D. Feature Selection using ANN

The feature selection method described in section IV.C for batch mode using ANN is applied on the feature vector $f = \{f_{[21]}, f_A, f_D, f_{[22]}\}$ to understand the more important features against the least important. For this experiment, we initialize attenuation function $F(M_i)$ to 0.05. After the ANN training using the feature selection technique, the values of $F(M_i)$ are shown in Figure 10 for all the 46 features. The first 18 features ($f_{[21]}$) are the one reported in [21], the next two features are based on area (f_A), the next 12 features are based on centroid distances (f_D) and the final 14 features ($f_{[22]}$) are as reported in [22]. It can be seen that most of the static features reported in [22] are having much higher importance than the dynamic features. The skeleton data captured using Microsoft SDK is not noise free and as the dynamic feature captures the property of persons walking in time, the uniformity in feature generation could be lost due to noise incorporation during walk, e.g different angle could be different in different gait cycles. On the other hand the static features capture the relative distances between joints and more stable in computation. Given the raw skeleton information using Microsoft SDK and Neural network as a classifier, these static features get more importance in person identification.

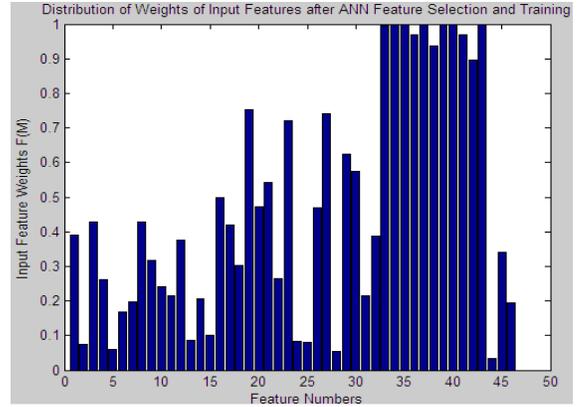


Figure 10. The value of attenuation function $F(M)$ for different features $f = \{f_{[21]}, f_A, f_D, f_{[22]}\}$

The recognition accuracies for 10 subjects with and without feature selection technique are given in TABLE V. It can be seen that with simultaneous feature selection and classification the F-score has increased to 0.62.

TABLE V. PERFORMANCE (F-SCORE IN %) COMPARISON WITH AND WITHOUT FEATURE SELECTION FOR 10 SUBJECTS

ANN without feature selection having features as $f = \{f_{[21]}, f_A, f_D, f_{[22]}\}$	ANN with feature selection having features as $f = \{f_{[21]}, f_A, f_D, f_{[22]}\}$
53	62

VI. CONCLUSION

In this paper we have presented few static and hybrid model based features for gait recognition solution using Microsoft Kinect. The static feature proposed is the area encompassed by the upper body and the hybrid features include the area encompassed by the portion below hip and the distances between the body-centroid and the centroids derived from the joints of the upper and lower limbs. The accuracy of the gait recognition is compared with the methods proposed in [21] and [22]. ANN based connectionist system [30] is used to perform simultaneous feature selection and classification in the batch mode. Results indicate that the static features get higher importance compared to the dynamic features. Moreover, the hybrid features proposed in this paper are having a great importance compared to the dynamic features proposed earlier in the ANN based feature selection process. Future scope of research lies in removing the noise from the skeleton data before extracting dynamic features and also investigates their importance in the feature selection process. Future research is also aimed to use other sources of human identification along with skeleton data from Kinect, to further improve the recognition accuracy. Moreover the effect of using multiple Kinect and camera needs to be analyzed in future.

ACKNOWLEDGMENT

We would like to acknowledge the colleagues of our Cyberphysical Systems Innovation Lab, Tata Consultancy

Services, who have helped us in data capturing activity in spite of their busy schedule in project work.

REFERENCES

- [1]. C. Liying, S. Qi, S. Han, C. Yang, Z. Shuying, "Design and implementation of human-robot interactive demonstration system based on Kinect", 24th Chinese Control and Decision Conference (CCDC), May 2012, pp.971-975.
- [2]. Arras, K.O., Lau, B., Grzonka, S., Lubner, M., Mozos, O.M., Meyer-Delius, D., Burgard, W. "Range-Based People Detection and Tracking for Socially Enabled Service Robots", Prassler, E., et al. (eds.) Towards Service Robots for Everyday Environ. STAR, vol. 76, pp. 235–280. Springer, Heidelberg 2012.
- [3]. H. Uwe, H. Sebastian Hommel, B. Michael, and D. Michael, "Face Detection and Person Identification on Mobile Platforms", E. Prassler et al. (Eds.): Towards Service Robots for Everyday Environ. STAR 76, pp. 227–234. Springer, Heidelberg 2012.
- [4]. <http://www.idiap.ch/scientific-research/themes/biometric-person-recognition>(last accessed on 9th Oct 2012)
- [5]. X. Qinghan, "Technology review - Biometrics-Technology, Application, Challenge, and Computational Intelligence Solutions", IEEE Computational Intelligence Magazine 2007, vol.2, pp. 5-25.
- [6]. D. Kim and J. Paik, "Gait recognition using active shape model and motion prediction", *IET Computer Vision* 2010, vol. 4, pp. 25-36.
- [7]. N. Borghese, L. Bianchi, and F. Lacquaniti, "Kinematic determinants of human locomotion" *J. Physiology* 1996,494(3):863.
- [8]. A. K. Jain, R. Bolle, and S. Pankanti, Biometrics: personal identification in networked society, Kluwer Academic Publishers 1999.
- [9]. Amit Kale, Aravind Sundaresan, A. N. Rajagopalan, Naresh P. Cuntoor, Amit K. Roy-Chowdhury, VolkerKrügerand Rama Chellappa, "Identification of Humans Using Gait", IEEE Transactions On Image Processing September 2004, vol. 13, no. 9.
- [10].R. Zhang, C. Vogler, and D. Metaxas, "Human gait recognition", IEEE Comp. Soc. Conf. On Computer Vision and Pattern Recognition Workshops 2004, pages 18–18.
- [11].Wang, Jin, She, Mary, Nahavandi, Saeid and Kouzani, Abbas, "A review of vision-based gait recognition methods for human identification", DICTA 2010: Proceedings of the Digital Image Computing: Techniques and Application, IEEE, Piscataway, N.J., pp. 320- 327.
- [12].C. BenAbdelkader, R. Cutler, and L. Davis, "Stride and cadence as a biometric in automatic person identification and verification", Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition 2002, pp. 372-377.
- [13].C. Yam, M. S. Nixon and J. N. Carter, "Automated person recognition by walking and running via model-based approaches", Pattern Recognition 2004, 37:1057-1072.
- [14].S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, "The humanIDgait challenge problem: data sets, performance and analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence 2005, vol. 27, pp. 162-177.
- [15].J. Han and B. Bhanu, 2, "Individual recognition using gait energy image", IEEE Transactions on Pattern Analysis and Machine Intelligence 2006, vol. 28, pp. 316-322.
- [16].A. F. Bobick and J. W. Davis, "The recognition of human movement using temporal templates", IEEE Transactions on Pattern Analysis and Machine Intelligence 2001, vol. 23, pp. 257-267.
- [17].L. Jianyi and Z. Nanning, 2007, "Gait History Image: A Novel Temporal Template for Gait Recognition", IEEE International Conference on Multimedia and Expo, 2007, pp. 663-666.
- [18].C. Chen, J. Liang, H. Zhao, H. Hu, and J. Tian,2009, "Frame difference energy image for gait recognition with incomplete silhouettes", Pattern Recognition Letters, vol. 30, pp. 977-984.
- [19].Z. Xue, D. Ming, W. Song, B. Wan, and S. Jin, "Infrared gait recognition based on wavelet transform and support vector machine," Pattern Recognition 2010, vol. 43, pp. 2904-2910.
- [20].W. Liang, T. Tieniu, N. Huazhong, and H. Weiming, "Silhouette analysis-based gait recognition for human identification," IEEE Transactions on Pattern Analysis and Machine Intelligence 2003, vol. 25, pp. 1505-1518.
- [21].Adrian Ball, David Rye, Fabio Ramos and Mari Velonaki, "Unsupervised Clustering of People from ‘Skeleton’ Data", HRI 2012, Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction, 2012.
- [22].Preis J., Kessel M., Werner M., Linnhoff-Popien C, "Gait Recognition with Kinect", 1stInternational Workshop on Kinect in Pervasive Computing, Newcastle, 2012.
- [23].S. Aravind, R. Amit, and C. Rama, "A hidden Markov model based framework for recognition of humans from gait sequences", Proceedings of the IEEE International Conference on Image Processing, 2003, pp. II-93-6 vol.3.
- [24].C. BenAbdelkader, R. Cutler, and L. Davis, "Motion-based recognition of people in EigenGait space", Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, 2002, pp. 267-272.
- [25].Y. Mu and D. Tao, "Biologically inspired feature manifold for gait, recognition", Neurocomputing 2010, vol. 73, pp. 895-902.
- [26].R. Hecht-Nielsen, Neurocomputing, Addison-Wesley: New York, Chapter 5, 1990.
- [27].D. E. Rumelhart, G. E. Hinton and R. J. Williams, "Learning internal representations by error propagation", Institute for Cognitive Science Report 8506, San Diego: University of California, 1985.
- [28].W. W. Hager and H. Zhang, "A new conjugate gradient method with guaranteed descent and an efficient line search", SIAM J. Optim. 2005, 16, pages 170-192.
- [29].M. Hagiwara, "A simple and effective method for removal of hidden units and weights", Neurocomputing, vol. 6, pp. 207-218, 1994.
- [30].N. R. Pal and K. K. Chintalapudi, "A connectionist system for feature selection", Neural Parallel Sci. Comput. 1997, vol. 5, no. 3, pp. 359–381.
- [31].Chakraborty, D., Pal, N.R.: Selecting Useful Groups of Features in a Connectionist Framework., IEEE Transactions on Neural Network (2008), pp. 381-396
- [32].E. Anderson, "The irises of the Gaspé Peninsula," Bull. Amer. IRIS Soc., vol. 59, pp. 2–5, 1935