Person Identification in Natural Static Postures Using Kinect

Vempada Ramu Reddy(✉), Kingshuk Chakravarty, and S. Aniruddha

Innovation Lab, Tata Consultancy Services Ltd., Kolkata, India
{ramu.vempada,kingshuk.chakravarty,aniruddha.s}@tcs.com

Abstract. Automatic person identification using un-obtrusive methods are of immense importance in the area of computer vision. Anthropometric approaches are robust to external factors including environmental illumination and obstructions due to hair, spectacles, hats or any other wearable. Recently, there have been efforts made on people identification using walking pattern of the skeleton data obtained from Kinect. In this paper we investigate the possibility of identification using static postures namely sitting and standing. Existing gait based identifications, mostly rely on the dynamics of the joints of the skeleton data. In case of static postures the motion information is not available, hence the identification mainly relies on the static distance information between the joints. Moreover, the variation of pose in a particular posture makes the identification more challenging. The proposed methodology, initially sub-divides the body-parts into static, dynamic and noisy parts followed by a combinatorial element responsible for selectively extracting features for each of those parts. Finally a radial basis function support vector machine classifier is used to perform the training and testing for the identification. Results indicate an identification accuracy of more than 97 % in terms of F-score for 10 people using a dataset created with various poses of natural sitting and standing posture.

Keywords: Person identification · Natural static posture · Skeleton joints · Kinect

1 Introduction

Human brain can discriminate between people based on their unique physical as well as behavioural characteristics [1]. Everyday the importance of non-intrusive person identification has been increasing as the technology that can serve several critical applications like video surveillance, people counting, server-room or datacenter authentication, audience measurement etc. Several modalities of person identification (PI) in terms of biometrics already exist in the current literature on computer vision. A few of them include behavioural characteristics like lip movement, typing pattern etc. or physiological signatures like speech,

L. Agapito et al. (Eds.): ECCV 2014 Workshops, Part II, LNCS 8926, pp. 793–808, 2015.
DOI: 10.1007/978-3-319-16181-5_60

face, iris, fingerprint etc. Unfortunately, these modalities are intrusive in nature, thus require direct human interaction for the authentication. Moreover extraction of fingerprint, iris or audio related biometric information (at recognizable form) from a large distance is definitely a challenging job. However, when other cues are not robust enough in discriminating between people, soft-biometrics like global shape [2] can be used to do the person identification. Global shape based approaches mainly utilize physical build of a person like body dimensions, height, length of limbs etc. for identifying a person. This type of systems is comparatively advantageous because it is very difficult to hide and conceal. In addition, global shape traits can also be extracted without making any user interaction, so it is non-intrusive in nature. They can be obtained either by using RGB-D images or by analysing skeleton joint co-ordinates of a particular subject. Fortunately, the Microsoft motion sensing device named Kinect directly provides RGB-D information and 3D co-ordinates of 20 skeleton joints like head, shoulder-center etc. In this paper, instead of storing image/video, we analyse structural build characteristics of a subject using only skeleton data which is more robust to illumination conditions. Skeleton joints can be obtained even if the face of the person is obstructed by hair, if person wear spectacles, hats or any other wearable. However, the skeleton joints obtained from Kinect is somewhat noisy only if the person wears black clothes which is mainly due to infrared sensor.

Several works have already been done on skeleton information based person identification using Kinect. Preis et al. [3] and Sinha et al. [4] did the same from side walking pattern using static as well as dynamic nature of gait features like length of arms, legs, velocity etc. Naresh et al. [5] had proposed a PI system from arbitrary unconstrained walking pattern. Though they [5] obtained 90% identification accuracy for 20 subjects, but the paths of the subjects were predefined (a front walking pattern with Kinect as the reference point) during training phase. Sinha et al. [6] investigated an interesting pose and subpose based concept for modeling arbitrary gait pattern using only skeleton data. They [6] employed unsupervised learning algorithm i.e., K-Means clustering, for identifying 3 poses and 8 subposes. Their method was able to achieve 94% recognition accuracy for 20 subjects. But, all of these skeleton based approaches aimed at identifying an individual based on only movement-pattern rather than static posture. Chakravarty et al. [7] proposed a PI system in static posture. Though they [7] got 96% identification accuracy for 10 subjects, but their method is mainly focused on frontal standing posture, rather than unconstrained natural static ones. In addition, they had carried out performance evaluation using training and testing at a fixed predefined position and posture. However as the subject is not very robotic and can assume variety of poses, their method performs very poorly in real-life. Identifying the person using global shape information obtained from RGB-D is quite easy compared to skeleton but it is quite challenging using skeleton data. For example, if two people are of same height and assume limb lengths are of similar size, still they can be easily discriminated from the width of hands, legs or body from RGB-D as it gives these additional clues. However, skeleton joints are single points we cannot get these crucial information like the

3D structure of person which is very unique. Therefore, identifying the persons of similar structures is quite challenging using skeleton data. Keeping all those problems in mind, we have developed a robust person identification system in static postures mainly sitting and/or standing using only global shape based features. In this work, we have defined sitting, standing, bending etc. as posture where a posture may have many poses. A pose is described as the attitude (e.g. orientation with respect to a reference point) of the body, or the position of the limbs (arms and legs) in a particular posture. While developing the robust PI system using skeleton data, we have explored physical build characteristics of a person in two phases where in the first phase we have explored feature sets related to constrained sitting and standing postures (method 1) and then, based on the drawback-analysis of method 1 in real-life scenario, method 2 is proposed in phase 2. The method 2 does not require any user cooperation and performs well in constrained as well natural sitting and standing postures. The contribution of this paper is mainly 4 folds

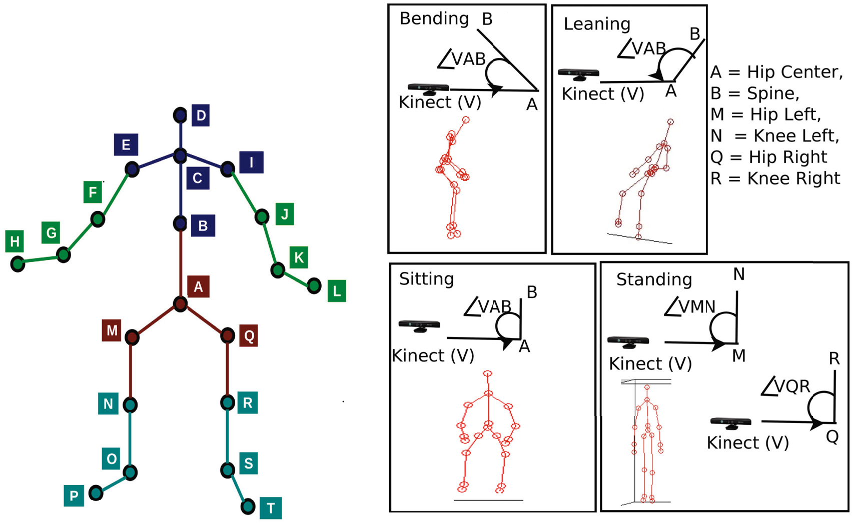
1. PI task is carried out in natural unconstrained static postures in real time using only skeleton data obtained from Kinect. The system is invariant to lighting condition and also ensures user's privacy.
2. Benefits and drawbacks of different feature sets are investigated for identifying an individual in natural and constrained static postures.
3. For robust PI, the pose invariant optimal feature vector is selected from different body-parts after examining combinatorial study on different feature sets.
4. Density based clustering approach is used for dividing entire skeleton structure based on static, dynamic and noisy nature of joints.

We have also evaluated performance of method 2 with respect to the state-of-the-art systems [6] [7] and it is shown that our method outperforms the existing systems in natural static postures.

Rest of the paper is organized as follows: Section 2 gives the brief explanation of posture and poses along with the details of database creation for static postures. Two phase implementation of our proposed PI system is presented in two sections Section 3 and Section 4 where Section 3 gives the performance analysis of different global shape based features on different datasets and Section 4 presents the proposed robust person identification system based on joint analysis of different body-parts. Conclusion of this paper is laid out in the final section.

2 Experimental Database

In this work, we have developed a person identification system using only skeleton information obtained from Kinect [8]. Here we are focusing on the PI task, only in static postures like sitting and standing. For this we have analyzed physical build characteristics in terms of skeleton data. Methods in [6] [4] [9] [3] [5] did the PI by analyzing the movement patterns in terms of spatio-temporal



(a) 20 skeleton joints with labels (b) Representation of postures using angles

Fig. 1. Representation of skeleton structure and postures

variation of skeleton joint co-ordinates. But, unfortunately no standard public database exists for person identification in static postures (specially sitting and standing) using skeleton data. Therefore, we have carefully designed our own database that suits to real-time scenario. In this study, we have used Kinect sensor which is placed at 6-10 ft distance from the subject to collect the skeleton data from sitting and standing posture. It mainly records $\{x, y, z\}$ co-ordinates (in meters) of different skeleton joints for a particular subject. The 20 skeleton joints namely Hip Center(A), Spine(B), Shoulder Center(C), Head(D), Shoulder Left(E), Elbow Left(F), Wrist Left(G), Hand Left(H), Shoulder Right (I), Elbow Right(J), Wrist Right(K), Hand Right(L), Hip Left(M), Knee Left(N), Ankle Left(O), Foot Left(P), Hip Right(Q), Knee Right(R), Ankle Right(S), Foot Right(T) obtained from the Kinect are shown in Fig. 1a. The data is collected from the sitting and standing postures in two modes - 1) *constrained static postures* - frontal sitting and standing pose, and 2) *unconstrained static postures* - natural sitting and standing pose. In both the modes, datasets are created from 10 people (3 female and 7 male). We have presented a brief discussion on posture and pose followed by the details on the corpus creation.

Overview of Posture and Pose

Before going into discussion about database creation on sitting and standing postures with different poses, we want to clarify the difference between posture and pose. Posture is viewed at macroscopic level whereas single posture can have multiple poses. The orientation of the posture with respect to some reference point is treated as pose. Therefore, pose can be viewed as containing microscopic level information. For example, the postures can be like sitting, standing, sleeping, bending, leaning etc. Any particular posture is independent of person's orientation in the space. However, pose should be defined with respect to some reference point. In our case, if we consider Kinect as the reference, then the orientation of person with respect to Kinect is treated as pose. If the subject is straight towards camera i.e., perpendicular it is treated as straight pose or frontal pose. Else we consider pose (with some angle with respect to Kinect) as natural one. Not only that in natural poses, the subject may vary position of his/her limbs. The skeleton joints extracted from Kinect are represented by 3D world co-ordinates (x,y,z) where 'x' represents the left/right variation, 'y' represents up/down variation and 'z' represents to/from variation of subject with respect to Kinect. Scientifically, the angles formed by some joints with respect to Kinect in X-Y (coronal) and/or Z-Y (sagittal) plane can differentiate the posture. Once posture is fixed, the orientation of subject with respect to Kinect in Z-X (transverse) plane can differentiate poses within the posture. As shown in Fig. 1b, postures like leaning, sitting, bending are defined using the angle information which is obtained from the joints like A and B (Fig. 1a) with respect to Kinect(marked as V). However, the posture standing is discriminated from other postures using additional angle information made by the joints M and N or Q and R (Fig. 1a).

Dataset #1
This is created from the respective 10 people in the constrained static postures specifically frontal sitting and standing one. In this case, we have asked the subjects to view straight towards the Kinect. From each subject, we have collected 1 set of data for training and 3 sets of data for testing. Each set consists of 1 minute of data with approximately 30 frames per second. The training set is frontal one where legs are kept perpendicular to Kinect and hands are lied on both the legs at different locations which are varied from Knee to near Hip location. One set of test data is similar to the training set whereas for other two test sets we have requested the subjects to remain in the frontal standing or sitting pose but asked to produce small variations of dynamic joints like leg and hand positions (without crossing legs and folding hands).

Dataset #2
This is also created from the same 10 people of dataset #1 but in unconstrained static poses i.e., natural standing and sitting poses. From each subject, we have collected 2 sets of data, where one set is used for training and other set is for testing. In the training phase, we have asked the subject to sit and stand in some particular predefined poses (one example shown in Fig. 2) but in testing phase we have not restricted the subject in viewing Kinect. Instead the subject is encouraged to sit and stand with some angle to the Kinect. In fact, we have requested the subject to give arbitrary sitting and standing posture by making large variations of dynamic joints. While designing dataset #2 we have emphasized the fact that in real life, during testing, a subject may give totally different static pose that is not present in the training corpus.

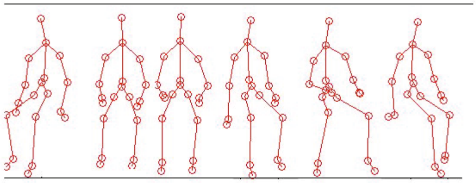


Fig. 2. Representation of different poses of sitting posture

3 Person Identification in Static Postures: Method 1

We have developed a person identification (PI) system from sitting and standing postures in two phases. In the first phase (method 1), the PI system is implemented in two steps (i) **feature extraction** (ii) **decision making and performance analysis**. The method 2 is proposed based on the performance analysis of method 1. In other words, we have critically analysed drawbacks of method 1 and proposed a robust PI system in phase 2 (method 2). It needs to be mentioned that the performance of method 1 & 2 is evaluated using both the datasets #1 & #2. The implementation details of method 1 are presented in the following subsections.

3.1 Feature Extraction

The feature extraction module generates different sets of features for identifying the person in sitting and standing posture. Therefore, identifying appropriate salient features from the 3D world co-ordinates of 20 joints, which can discriminate the individual characteristics, is a very crucial step for any high performance system. The details of features for PI are as follows:

In static postures, meaningful information about identity or uniqueness of any individual can be obtained by extracting the features related to the structural or physical build of the subject (e.g. height, length of limbs etc.). So keeping this fact in mind, we have used differences of 3D world co-ordinates between every pair of joints (physically connected and unconnected) as a candidate feature vector ($\boldsymbol{F}_{cu}$) and $\boldsymbol{F}_{cu}$ is extracted at frame level. The feature set $\boldsymbol{F}_{cu}$ contains all the necessary and unique information about the physical build of a subject whereas, differences of co-ordinate 'x', 'y' and 'z' for every joint-pair capture the width, height and depth information, respectively. From Fig. 1a it is observed that there are 20 joints with 19 physically connected pairs, where the differences of co-ordinates 'y' inherently give the information about length of limbs. The features $\boldsymbol{F}_c$ and $\boldsymbol{F}_y$ represent the differences of 3D world co-ordinates and differences of 'y' co-ordinate between every '*connected*' pair of joints, respectively. In the first phase of our implementation, the candidate features such as $\boldsymbol{F}_c$ and $\boldsymbol{F}_y$ ($\boldsymbol{F}_y$, $\boldsymbol{F}_c \subset \boldsymbol{F}_{cu}$) are extracted from each frame for analysing how they affect PI in different posing conditions. $\boldsymbol{F}_{cu}$, $\boldsymbol{F}_c$ and $\boldsymbol{F}_y$ are formulated using equations (1), (2) and (3) where J is the total number of joints in D dimensional co-ordinate system and CP represents number of physically connected joint-pairs.

$$\begin{aligned} &\boldsymbol{F}_{cu} = abs((x^j, y^j, z^j) - (x^k, y^k, z^k)) \ \forall\ j = [1, 20], k = [1, 20], j \neq k, \\ &\boldsymbol{F}_{cu} \in \mathbf{R}^{(\mathbf{D} \times {}^{\mathbf{J}}\mathbf{C_2})}, \text{where } \mathbf{D}{=}3 \text{ and } \mathbf{J}{=}20 \end{aligned} \tag{1}$$

$$\begin{aligned} &\boldsymbol{F}_{c} = abs((x^j, y^j, z^j) - (x^k, y^k, z^k)) \ \forall\ j, k = \{1, \ldots, 20 | j, k \text{ connected}\}, \\ &\boldsymbol{F}_{c} \in \mathbf{R}^{(\mathbf{D} \times \mathbf{CP})} \text{ and } \boldsymbol{F}_c \subset \boldsymbol{F}_{cu}, \text{where } \mathbf{D}{=}3 \text{ and } \mathbf{CP}{=}19 \end{aligned} \tag{2}$$

$$\begin{aligned} &\boldsymbol{F}_{y} = abs((y^j) - (y^k)),\ \forall\ j, k = \{1, \ldots, 20 | j, k \text{ connected}\}, \\ &\boldsymbol{F}_{y} \in \mathbf{R}^{\mathbf{CP}} \text{ and } \boldsymbol{F}_y \subset \boldsymbol{F}_{cu}, \text{where } \mathbf{CP}{=}19 \end{aligned} \tag{3}$$

3.2 Decision Making and Performance Analysis

The decision making task is carried out using a supervised learning algorithm with feature sets $\boldsymbol{F}_{cu}$, $\boldsymbol{F}_c$ and $\boldsymbol{F}_y$ separately. A classification algorithm is used to map feature vectors to a particular object class representing a person. We have realized the classifier using multi-class support vector machine (SVM) with Radial Basis Function (RBF) as kernel [10] [11]. SVM classification is an example of supervised learning. SVMs are useful due to their wide applicability for classification tasks in many applications [12]- [18]. The main goal of SVM for classification problem is to produce a model which predicts target class label of data instances in the testing set, given only the attributes. The intuition to use RBF kernel function is due to its universal approximation properties. Also, it offers good generalization as well as good performance in solving practical problems [15] [16].

In this study, the statistical measure F-score [6], which is defined as the harmonic mean of precision and recall is used for performance evaluation. For N subjects F-Score is defined by the equation (4).

$$\text{F-score}_i = \frac{2 * precision_i * recall_i}{(precision_i + recall_i)} \quad \forall i, 1 \leq i \leq N \tag{4}$$

For method 1, various types of experiments are then carried out on the datasets explained in the section 2. These are described as follows:

(A) Trained and Tested at Frontal Static Posture.
As an initial step of our experimentation, we have used only dataset #1 for PI in frontal static posture. The identification accuracy in the form of confusion matrix for test set 1 using feature vector $\boldsymbol{F}_{cu}$ is given in Table 1. Table 2 represents the average F-scores of the PI system using feature vectors $\boldsymbol{F}_{cu}$, $\boldsymbol{F}_c$ and $\boldsymbol{F}_y$ separately on all the 3 test sets.

Analysis: The average performance of PI system shown in the diagonal of Table 1 indicates that almost all persons are well classified. But it is also observed from Table 2 that for all the features, performance of method 1 is better on test set

Table 1. Confusion matrix for 10 subjects trained and tested at frontal sitting posture using feature vector F_{cu}. Entries in table indicate F-scores in (%)

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}
P_1	**100.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
P_2	1.85	**98.15**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
P_3	0.00	0.00	**100.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00
P_4	0.00	0.00	0.00	**100.00**	0.00	0.00	0.00	0.00	0.00	0.00
P_5	0.00	0.00	0.00	0.00	**97.23**	2.01	0.76	0.00	0.00	0.00
P_6	0.00	0.00	0.00	0.00	0.00	**100.00**	0.00	0.00	0.00	0.00
P_7	0.00	0.00	0.00	0.00	0.00	0.00	**100.00**	0.00	0.00	0.00
P_8	0.00	0.65	0.00	0.00	0.00	0.00	0.00	**99.35**	0.00	0.00
P_9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**100.00**	0.00
P_{10}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**100.00**

Table 2. Average F-scores(%) of PI system for frontal training and testing

Test on dataset #1	**Sitting**			**Standing**		
	$\boldsymbol{F}_{cu}$	$\boldsymbol{F}_{c}$	$\boldsymbol{F}_{y}$	$\boldsymbol{F}_{cu}$	$\boldsymbol{F}_{c}$	$\boldsymbol{F}_{y}$
Set 1	99.47	98.61	97.79	100.00	99.42	98.18
Set 2	92.10	96.78	94.56	91.27	95.65	92.18
Set 3	87.45	90.19	89.20	90.13	92.90	92.14

Table 3. Average F-score(%) of PI system for frontal training and natural testing

Test on dataset #2	$\boldsymbol{F}_{cu}$	$\boldsymbol{F}_{c}$	$\boldsymbol{F}_{y}$
Natural Sit	54.60	62.17	58.19
Natural Stand	63.22	68.98	65.37

1 compared to test sets 2 and 3. This is mainly because the test set 1 and the training set have similar poses for the postures. However, if the subject even slightly varies his/her frontal sitting or standing pose (dataset #1–> test sets 2 & 3) like keeping the arm and leg positions different from that of training model, the performance of this implementation degrades (F-scores for set 2 and set 3 in Table 2). Moreover, as $\boldsymbol{F}_{cu}$ includes differences of 3D co-ordinates for both connected and unconnected pairs, it is obvious that $\boldsymbol{F}_{c}$ and $\boldsymbol{F}_{y}$ perform relatively better on test sets 2 & 3 than $\boldsymbol{F}_{cu}$. Therefore, slight variation in legs and arm positions in testing phase largely affects feature vectors related to the unconnected joint-pairs which is present $\boldsymbol{F}_{cu}$.

(B) Trained at Frontal and Tested Using Unconstrained Static Posture. To make our PI system more realistic, we have used frontal sitting and standing data from dataset #1 for training and unconstrained (natural) pose data from dataset #2 for testing. The average F-scores for all feature vectors are compared in Table 3.

Analysis: Table 3 clearly tells us that the results are more worse compared to Table 2. Our analysis suggests that the system performs poorly because of lack of pose variation information in the training data.

(C) Trained and tested at natural static posture
Next, both the training and testing data are taken from dataset #2. The diagonal entries in Table 4 show the average PI performance for 10 subjects using feature vector $\boldsymbol{F}_{cu}$. We have also compared the performance using feature vectors $\boldsymbol{F}_{cu}$, $\boldsymbol{F}_c$ and $\boldsymbol{F}_y$ in Table 5.

Table 4. Confusion matrix for 10 subjects trained and tested at natural sitting posture using the feature vector F_{cu}. Entries in table indicate F-scores in (%)

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}
P_1	**62.07**	13.12	0.00	0.00	0.00	14.81	0.00	4.48	0.00	5.52
P_2	6.35	**91.29**	0.00	0.05	2.31	0.00	0.00	0.00	0.00	0.00
P_3	21.48	0.00	**52.10**	6.36	0.00	0.00	0.00	5.62	13.18	1.26
P_4	1.83	0.00	6.67	**89.58**	0.00	0.00	1.92	0.00	0.00	0.00
P_5	20.40	0.00	0.00	0.00	**61.16**	0.00	3.60	0.00	14.84	0.00
P_6	0.00	2.62	20.10	0.00	0.00	**75.23**	0.00	0.00	2.05	0.00
P_7	0.00	0.00	45.17	0.00	0.00	0.00	**54.83**	0.00	0.00	0.00
P_8	0.00	9.22	2.67	16.46	0.00	0.00	0.00	**71.65**	0.00	0.00
P_9	0.00	0.06	28.70	0.00	0.00	2.80	0.85	0.00	**67.59**	0.00
P_{10}	0.00	0.00	2.63	0.98	15.28	0.06	0.00	0.08	0.00	**80.97**

Table 5. Average F-score(%) of PI system for natural training and testing

Test on dataset #2	$\boldsymbol{F}_{cu}$	$\boldsymbol{F}_c$	$\boldsymbol{F}_y$
Natural Sit	70.65	69.80	65.57
Natural Stand	79.12	73.29	69.65

Analysis: From Tables 4 and 5, it is observed that the average performance of method 1 is slightly improved compared to the previous approach. However, the performance of PI is still not satisfactory and we have got maximum 70.65% and 79.12% PI accuracies in natural sitting and standing postures, respectively. Our analysis suggests that even the subject maintains different pose but may not have good control on hands and leg positions due to flexibility of more dynamic nature of joints in natural scenario. It is also seen that some of the joints exhibit noise in some viewing angles due to occlusion and thus make the PI system more erroneous. From the above analysis, we conclude that different features perform better in different conditions for method 1. Hence, if we can carefully select the features based on the orientation of joints, it will definitely improve the system performance. This gives us the motivation to develop more robust PI system by modifying method 1. The following section 4 describes our modified approach.

4 Person Identification in Static Postures: Method 2

We always keep in mind that we have to design a PI system in natural static posture so that it perfectly matches any real-life scenario. Therefore we have developed method 2 by modifying method 1 to overcome the above limitations (described in the section 3). In method 2, we have analyzed joints belong to different body-parts, extracted relevant features and then finally evaluated the performance. The frame-work of method 2 contains 5 modules (i) **Data Acquisition (DA)**, (ii) **Skeleton Divider (SD)**, (iii) **Feature Generator (FG)**, (iv) **Combinatorics Engine (CE)** and (v) **Model Generator (MG)** in its functional architecture which is shown in Fig. 3.

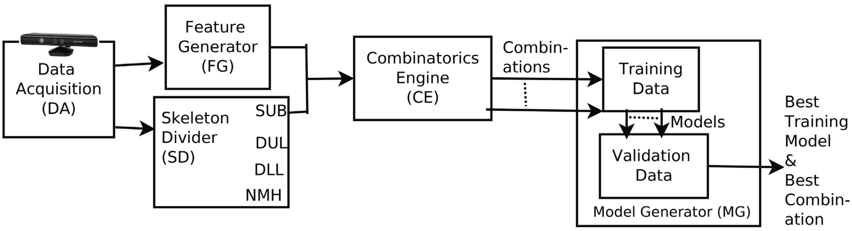


Fig. 3. Functional architecture of method 2

The DA module captures co-ordinates of 20 skeleton-joints using Kinect and forwards these 20 joints to SD module. In SD module, entire skeleton structure is divided into different body-parts based on the static upper, dynamic lower, dynamic upper and noisy middle nature of joints (labeled with different colors in Fig. 1a). FG module extracts the candidate features $\boldsymbol{F}_{cu}$, $\boldsymbol{F}_c$ and $\boldsymbol{F}_y$ (explained in the section 3) from the skeleton joints. Once the feature generation is done, CE module explores all possible combination of features extracted from different body-parts and finally, all these combinations are feeded to MG module to generate the training models. In method 2, the training and testing are done only on dataset #2 where predefined poses are used for training but testing is carried out with unconstrained natural static postures. The key contribution of this approach is mainly dividing 20 skeleton joints into different body-parts and automatic selection of optimal features from the respective body-parts. In addition to this we have also analyzed the influence of certain angles in capturing the pose related information. The details of the proposed methodology and influence of the angles on the proposed system are presented in the following subsections.

4.1 Methodology

In the static posture like sitting or standing, a person can be oriented in any direction with respect to Kinect exhibiting natural pose. However, for a given posture a person can not move some of the joints flexibly irrespective of poses.

For example, upper body joints like Spine, Hip Center etc. are fixed for any pose in a particular posture. We define those joints as static one. On the contrary, in a single pose, a subject can move his joints like Knee Left, Wrist Left, Foot Left etc. very flexibly. Therefore, we name them as dynamic joints. It is also noticed that some of the joints are more prone to noise due to occlusion effect. For example, in most of the poses of sitting posture, Hip Left and Hip Right are occluded with Knee Left and Knee Right, respectively. These types of joints are considered as noisy joints. This is also verified by grouping the co-ordinates of different joints from upper and lower body-parts using density based clustering algorithm DBSCAN [19]. The results of DBSCAN for some joints (Left portion of the body) for both the postures are illustrated in Table 6. Right portion of the body joints also exhibited the similar trend. Table 6 indicates that for sitting posture(s) DBSCAN identifies 6 clusters whereas for standing posture(s) it forms only 2 clusters. It can also be noticed from the results that in both postures, certain joints of upper body like Shoulder Center, Shoulder Left, Spine etc. form one cluster (static cluster) and Elbow Left, Wrist Left, Knee Left, Ankle Left form another cluster (dynamic cluster). The joints which are varying over frames mainly belong to dynamic cluster and the joints which are static over frames form the static cluster. However, Table 6 also captures an interesting fact that for Hip portion joints like Hip Left, Hip Center, the frames are not clearly separable as pure static or dynamic ones because Hip portion joints are occluded by Knee portion joints while sitting, this causes the noisy nature of Hip joints. 35.02% and 16.09% of HipLeft frames (Table 6) are moved to dynamic cluster in sitting and standing posture. This is mainly because in sitting, the occlusion is more compared to standing. The dynamic joints are further divided into two portions namely dynamic upper and dynamic lower. So, based on the above observation, entire skeleton structure is divided into four parts:

1. *Static Upper Body (SUB)*: The joints B, C, D, E and I representing the main body portion (color coded in dark blue in Fig. 1a) are more static in nature during any pose for a particular posture.
2. *Dynamic Upper Limbs (DUL)*: Based on the subject's flexibility of changing the arm positions in natural static postures, the joints F, G, H, J, K and L are considered as dynamic upper limbs (color coded in green in Fig. 1a).
3. *Dynamic Lower Limbs (DLL)*: Based on the subject's flexibility of changing leg positions in natural static postures, we have considered the joints N, O, P, R, S and T as dynamic lower limbs (color coded in sky blue in Fig. 1a).
4. *Noisy Middle Hip (NMH)*: It is also noticed that if the person varies his/her pose in a particular posture, some joints are reliable and some are noisy. It is mainly due to occlusion of some body portions. This effect is very much vivid in middle hip portion. Therefore, we name the joints A, M and Q as noisy middle hip joints (color coded in deep red in Fig. 1a).

When the body-part segmentation is done, we have explored all the possible combination of features ($\boldsymbol{F}_{cu}$, $\boldsymbol{F}_c$ and $\boldsymbol{F}_y$) extracted from those body-parts

Table 6. Division of body parts using clustering of joints where Cl.=Cluster

Sitting						Joint	Standing	
Cl. 1	Cl. 2	Cl. 3	Cl. 4	Cl. 5	Cl. 6		Cl. 1	Cl. 2
29.96	0.67	*64.66*	0.67	0.00	4.04	HipCenter	*18.15*	*77.85*
4.71	0.54	**93.67**	0.40	0.00	0.67	Spine	13.00	**87.00**
3.10	0.27	**95.69**	0.40	0.13	0.40	ShoulderCenter	13.34	**86.66**
12.11	0.00	**87.08**	0.40	0.13	0.27	ShoulderLeft	4.26	**95.74**
88.83	2.56	3.90	0.67	0.67	3.36	ElbowLeft	**91.20**	8.80
97.98	0.40	0.40	0.00	0.27	0.94	WristLeft	**91.20**	8.80
94.75	0.54	2.83	1.21	0.13	0.54	KneeLeft	**89.96**	10.04
99.33	0.13	0.13	0.00	0.27	0.13	AnkleLeft	**88.72**	11.28
35.02	1.08	*60.94*	0.27	0.54	2.15	HipLeft	*16.09*	*83.91*

(SUB, DUL, DLL and NMH). This is carried out by CE module and it generates total number of combinations $= \sum_{k=1}^{TP} p^k \times \binom{TP}{k}$, where $TP=$ total number of body-parts and $p =$ total number of features. With 3 type of features and 4 body parts, different features extracted from single body part result to 12 combinations $(3^1 \times \binom{4}{1})$. For example, if 3 feature vectors is extracted from single body part at a time and no features are extracted from other body parts, it can be done in three ways. In the same way, three feature vectors extracted from rest of the body parts can be done in 9 ways. Therefore, features extracted from single body part scheme results to total 12 combinations. Similarly, feature vectors extracted from two body parts at a time while maintaining other two body-parts features none can result 54 combinations $(3^2 \times \binom{4}{2})$. Three feature vector combinations for 3 different body parts and no features from left body part will result 108 combinations $(3^3 \times \binom{4}{3})$. Finally, different feature vector combinations including all the body parts result 81 $(3^4 \times \binom{4}{4})$. Thus the system has result to 255 combinations in total. Then all these combinations are feeded to multi-class SVM to generate different models. Now to do the evaluation, we have done 5 fold cross-validation using the training corpus from dataset #2. The average of top 10 PI F-scores are listed in Table 7 for both sitting and standing postures. In Table 7, 'NOT' indicates that none of the feature vectors are employed for that particular body-part. It is found that among all the 255 models, the combination (F_{sit}^{best})– $\boldsymbol{F}_{cu}$ for SUB, $\boldsymbol{F}_c$ for DUL, $\boldsymbol{F}_y$ for DLL and 'NOT' for NMH produces the best F-Score in sitting posture. Similarly for standing posture, we compute the same i.e. F_{stand}^{best}. To test the robustness of method 2, these combinations are applied on the test data of dataset #2 and we able to achieve average 93.00% & 95.33% identification accuracy in sitting and standing, respectively. Table 8 shows the confusion matrix in natural sitting posture for the combination F_{sit}^{best}.

Table 7. Top 10 F-scores(%) of method 2 using combination of features and bodyparts (Cross-validation performance)

Sl. No.	Sitting					Standing				
	SUB	DUL	DLL	NMH	F-score (%)	SUB	DUL	DLL	NMH	F-score(%)
1	$\boldsymbol{F}_{cu}$	$\boldsymbol{F}_c$	$\boldsymbol{F}_y$	NOT	95.51	$\boldsymbol{F}_{cu}$	$\boldsymbol{F}_c$	$\boldsymbol{F}_y$	$\boldsymbol{F}_{cu}$	96.02
2	$\boldsymbol{F}_{cu}$	NOT	$\boldsymbol{F}_y$	NOT	94.98	$\boldsymbol{F}_{cu}$	NOT	$\boldsymbol{F}_c$	NOT	95.73
3	$\boldsymbol{F}_{cu}$	NOT	$\boldsymbol{F}_y$	$\boldsymbol{F}_{cu}$	93.01	$\boldsymbol{F}_{cu}$	$\boldsymbol{F}_c$	$\boldsymbol{F}_y$	NOT	94.56
4	$\boldsymbol{F}_{cu}$	$\boldsymbol{F}_y$	$\boldsymbol{F}_y$	NOT	91.71	$\boldsymbol{F}_{cu}$	NOT	$\boldsymbol{F}_y$	$\boldsymbol{F}_{cu}$	94.05
5	$\boldsymbol{F}_{cu}$	$\boldsymbol{F}_y$	$\boldsymbol{F}_y$	$\boldsymbol{F}_{cu}$	91.27	$\boldsymbol{F}_{cu}$	NOT	$\boldsymbol{F}_y$	NOT	93.72
6	$\boldsymbol{F}_{cu}$	$\boldsymbol{F}_c$	$\boldsymbol{F}_c$	$\boldsymbol{F}_y$	90.86	$\boldsymbol{F}_{cu}$	$\boldsymbol{F}_c$	$\boldsymbol{F}_y$	$\boldsymbol{F}_y$	91.00
7	$\boldsymbol{F}_{cu}$	$\boldsymbol{F}_c$	NOT	$\boldsymbol{F}_{cu}$	89.98	$\boldsymbol{F}_{cu}$	$\boldsymbol{F}_y$	$\boldsymbol{F}_c$	$\boldsymbol{F}_{cu}$	90.72
8	$\boldsymbol{F}_c$	NOT	$\boldsymbol{F}_c$	$\boldsymbol{F}_y$	89.45	$\boldsymbol{F}_{cu}$	$\boldsymbol{F}_c$	$\boldsymbol{F}_c$	$\boldsymbol{F}_y$	90.57
9	$\boldsymbol{F}_{cu}$	$\boldsymbol{F}_c$	NOT	NOT	89.38	$\boldsymbol{F}_{cu}$	$\boldsymbol{F}_c$	$\boldsymbol{F}_c$	NOT	90.57
10	$\boldsymbol{F}_{cu}$	$\boldsymbol{F}_c$	$\boldsymbol{F}_c$	$\boldsymbol{F}_{cu}$	89.29	$\boldsymbol{F}_{cu}$	$\boldsymbol{F}_y$	$\boldsymbol{F}_y$	$\boldsymbol{F}_{cu}$	90.36

Table 8. Confusion matrix for 10 subjects using F_{sit}^{best} on test-set of dataset #2

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}
P_1	**94.21**	0.00	1.21	2.43	0.00	2.15	0.00	0.00	0.00	0.00
P_2	0.00	**89.79**	0.00	6.14	2.00	0.00	0.00	2.07	0.00	0.00
P_3	3.33	0.00	**90.18**	0.00	3.16	0.00	0.00	0.00	0.00	3.33
P_4	0.00	0.00	0.00	**96.67**	0.00	0.00	0.00	3.33	0.00	0.00
P_5	0.00	6.66	3.33	0.00	**88.02**	0.00	0.00	0.00	1.99	0.00
P_6	2.00	0.00	0.00	0.00	0.00	**98.00**	0.00	0.00	0.00	0.00
P_7	0.00	0.00	0.00	0.00	5.02	0.00	**92.50**	0.00	0.38	2.10
P_8	0.00	0.00	0.00	5.33	0.00	0.00	0.00	**94.67**	0.00	0.00
P_9	1.50	0.00	0.00	6.67	0.00	4.02	0.00	0.00	**87.41**	0.40
P_{10}	0.00	0.00	1.50	0.00	0.00	0.00	0.00	0.00	0.00	**98.50**

Analysis: The top 10 results indicate that in many cases if the features extracted from the body-part NMH are not considered then the performance is better. Even best PI accuracy in sitting posture is obtained without using NMH joints (Row 1 Table 7). It is observed in Table 7 that 'NOT' for NMH joints appeared four times. However, it is observed from the results of 255 combinations this effect is less in standing posture due to less occlusion of NMH joints. Due to space constraint we have not given all 255 combinations. In some cases, it is also seen that if we do not use features from DUL and DLL, method 2 provides good results. This analysis helps us to conclude that the joints belong to NMH are more noisy than the others. Moreover, some joints of DUL and DLL produce noise when the person sits or stands with some orientation other than frontal pose. It is mainly because of occlusion of joints by other body-parts. Therefore, some frames get misclassified which results in slightly reduced performance. Table 7 also emphasizes that in all top 10 results most of the times, DLL and DUL use features $\boldsymbol{F}_y$

and $\boldsymbol{F}_c$ but not $\boldsymbol{F}_{cu}$. It is because, the features related to connected joint pairs are sufficient enough to capture the dynamic nature of joints specially arms and legs. Similarly, $\boldsymbol{F}_{cu}$ captures all the information related to static nature of upper main body portion. As the joints belonging to the SUB part are more static in nature, the body segment is proved to be most stable one across all poses in any postures. Not only that, we also explore different angles made by SUB-joints to capture the variation of poses in any natural static posture.

4.2 Influence of Angles

If the person sits or stands in natural pose, the orientation of main body is very crucial for defining a pose. It can be easily captured by computing angles formed by the joints C, E and I from shoulder portion and A, M and Q from hip portion (with respect to Kinect (V)) in Z-X plane. Table 9 shows the effect of these four angles namely ∠VCE, ∠VCI, ∠VAM and ∠VAQ on PI system.

Table 9. F-scores(%) of method 2 without and with angles and F-scores(%) with the methods proposed in [6] & [7]. In sitting $F^{best} = F^{best}_{sit}$ & in standing $F^{best} = F^{best}_{stand}$, and V is Kinect position

Posture	F^{best}	F^{best}, ∠VCE, ∠VCI, ∠VAM and ∠VAQ	F^{best}, ∠VCE and ∠VCI	[6]	[7]
Sitting	93.00	*89.63*	**96.81**	11.16	15.29
Standing	95.33	*93.61*	**97.65**	31.28	20.73

Analysis: Table 9 clearly shows that in both static postures, the performance of method 2 is degraded with the inclusion of these four angles along with the optimal combination F^{best}_{sit} & F^{best}_{stand} (shown in italics column 3 in Table 9). This is mainly due to the inclusion of angles formed by more noisy joints like A, M and Q. However the degradation in performance is less in standing posture compared to sitting one as the occlusion of hip portion is less in natural standing. After removal of these angles (∠VAM and ∠VAQ), it is observed that the performance of method 2 is enhanced further compared to previous one (shown in column 2 and 4 in Table 9). From this we infer that angles formed by the shoulder joints are the key contributors in capturing the variation of pose information in the natural unconstrained static postures. In this study, we have done the step-by-step analysis for making the PI system in static postures more robust and realistic. It needs to be mentioned that using method 2, the feature set F^{best}, ∠VCE and ∠VCI is able to achieve average 96.81% and 97.65% identification accuracy, in sitting and standing postures, respectively.

For the sake of completion of the analysis, we compare with the features proposed earlier for walking pattern in [6]. Sinha et al. had done the pose and subpose based modeling using static and dynamic gait features in [6]. We have also tested their approach on our dataset #2. But their performance on our dataset is not very satisfactory. It is mainly because their proposed features related to poses and subposes [6] fail to model pose variations in static postures. In addition, we have explored the method mentioned by Chakravarty et al. [7] on dataset #2. As the feature vector used in [7], is strictly focused on constrained frontal standing pose, their system fails to identify most of the subjects in natural sitting and standing poses. The performance comparison of method 2 with the state-of-the-art systems [6] & [7] is presented in the last 2 columns of Table 9. As expected the features for walking or constrained standing posture are not good for the unconstrained natural static scenario.

5 Conclusions

In this work, we have proposed a PI system in 2 phases. In the first phase, different sets of global shape based features which represent the identity of the person are explored. These features are then extracted from constrained and unconstrained datasets of sitting and standing postures. Based on the analysis and drawbacks of certain features for different body-parts in different poses, robust PI system is proposed in phase 2. In phase 2, clustering algorithm is used to identify static, dynamic and noisy joints. From that analysis, entire skeleton body is divided into four segments and we have explored all possible combinations of features from these segments. It greatly improves PI accuracy from 70.65% to 93% in sitting and 79.12% to 95.33% in standing posture. The effect of angle information from shoulder and hip portions is also analysed and it is found that inclusion of angles from hip portion degrades the system performance whereas angles extracted from shoulder portion enhances PI accuracy to 96.81% and 97.65% for both sitting and standing postures, respectively. Performance evaluation matrices also portray the significant improvement of identification accuracy in static postures over the contemporary systems. In future, we like to incorporate more static postures in our proposed system. We have also like to improve the system performance accuracy use angle information obtained from different joints i.e, transforming all poses to frontal pose using angle information and then extracted the features. Moreover we have a plan to combine our approach with other soft-biometric traits like gait, skin color etc. to build a multimodal PI system.

References

1. Anastassiou, G.A., Duman, O.: Introduction. In: Anastassiou, G.A., Duman, O. (eds.) Towards Intelligent Modeling: Statistical Approximation Theory. ISRL, vol. 14, pp. 1–8. Springer, Heidelberg (2011)
2. Jain, A.K., Dass, S.C., Nandakumar, K.: Can soft biometric traits assist user recognition? In: Defense and Security, International Society for Optics and Photonics, pp. 561–572 (2004)
3. Preis, J., Kessel, M., Werner, M., Linnhoff-Popien, C.: Gait recognition with kinect. In: 1st International Workshop on Kinect in Pervasive Computing (2012)

4. Sinha, A., Chakravarty, K., Bhowmick, B.: Person identification using skeleton information from kinect. In: ACHI 2013, The Sixth International Conference on Advances in Computer-Human Interactions, pp. 101–108 (2013)
5. Kumar, M., Babu, R.V.: Human gait recognition using depth camera: a covariance based approach. In: Proc. of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP), vol. 20. ACM (2012)
6. Sinha, A., Chakravarty, K.: Pose based person identification using kinect. In: IEEE International Conference on Systems, Man, and Cybernetics (SMC) 2013, pp. 497–503 (2013)
7. Chakravarty, K., Chattopadhyay, T.: Frontal-standing pose based person identification using kinect. In: Kurosu, M. (ed.) HCI 2014, Part II. LNCS, vol. 8511, pp. 215–223. Springer, Heidelberg (2014)
8. Microsoft: Kinect sdk (2012). http://www.microsoft.com/en-us/kinectforwindows/develop/developer-downloads.aspx. Accessed 29 June 2014
9. Ball, A., Rye, D., Ramos, F., Velonaki, M.: Unsupervised clustering of people from skeletondata. In: 2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 225–226. IEEE (2012)
10. Cortes, C., Vapnik, V.: Support-vector networks, vol. 20, pp. 273–297. Springer (1995)
11. Dietterich, T.G., Bakiri, G.: Solving multiclass learning problems via error-correcting output codes (1995)
12. Koolagudi, S.G., Reddy, R., Rao, K.S.: Emotion recognition from speech signal using epoch parameters. In: 2010 International Conference on Signal Processing and Communications (SPCOM), pp. 1–5. IEEE (2010)
13. Rao, K.S., Reddy, R., Maity, S., Koolagudi, S.G.: Characterization of emotions using the dynamics of prosodic. Proc. speech prosody, vol. 4 (2010)
14. Rao, K.S., Koolagudi, S.G., Vempada, R.R.: Emotion recognition from speech using global and local prosodic features. International Journal of Speech Technology **16**(2), 143–160 (2013)
15. Reddy, V.R., Sinha, A., Seshadri, G.: Fusion of spectral and time domain features for crowd noise classification system. In: 2013 13th International Conference on Intelligent Systems Design and Applications (ISDA), pp. 1–6. IEEE (2013)
16. Reddy, V.R., Chattopadhyay, T.: Human activity recognition from kinect captured data using stick model. In: Kurosu, M. (ed.) HCI 2014, Part II. LNCS, vol. 8511, pp. 305–315. Springer, Heidelberg (2014)
17. Vempada, R., Kumar, B., Rao, K.: Characterization of infant cries using spectral and prosodic features. In: 2012 National Conference on Communications (NCC), pp. 1–5. IEEE (2012)
18. Chattopadhyay, T., Reddy, V.R., Garain, U.: Automatic selection of binarization method for robust ocr. In: 2013 12th International Conference on Document Analysis and Recognition (ICDAR), pp. 1170–1174. IEEE (2013)
19. Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. Kdd. **96**, 226–231 (1996)