

Person Recognition Based on Face and Body Information for Domestic Service Robots

Shaghayegh Gharghabi

Robotic Research Center

Amirkabir University of Technology (Tehran Polytechnic)

Tehran, Iran

s.gharghabi@aut.ac.ir

Reza Safabakhsh

Department of Computer Engineering

Amirkabir University of Technology (Tehran Polytechnic)

Tehran, Iran

safa@aut.ac.ir

Abstract—Non-intrusive person recognition in different poses is a crucial task in the field of human-robot interaction (HRI). Most of the previous work for person recognition were based on face. In these approaches, if the person's face is invisible to the robot, it will ask the person to stand in a location suitable for the robot and look at the robot to recognize him. However in a normal interaction, the robot should be able to recognize a person without user cooperation to have acceptable interaction with users in real-time. In this paper, we address this problem using a combination of face and 3D information of a person's body. We used a mathematical framework based on the Bayesian decision theory for decision making. This combination can be used for recognizing scenarios where the pose of the people is unconstrained which makes it difficult to recognize a person by applying common face recognition algorithms. We tested our proposed approach in recognizing a five person group in various poses such as sitting and circular walking. The method was applied to a service robot equipped with the Kinect sensor. The results show a mean accuracy of 40.48% over all the poses using face alone, an accuracy of 67.23% using body alone, and an accuracy of 80.23% using the combination of the face and body information.

Keywords— *Person Recognition; Service robot; Human Robot Interaction; Kinect; Face recognition.*

I. INTRODUCTION

Over the past years, a number of service robots have been developed which work in indoor environments such as hospitals, museums, office buildings and supermarkets [1]. Daily-life assistance in domestic environments is an upcoming application of service robots. In this context, personal service robots, known as domestic robots, will soon become an essential part of modern life [2].

Domestic robots are designed to perform tasks such as delivery, entertainment, cleaning, washing, surveillance and ironing [3]. One of the most important challenges to be solved in this field is an unconstrained and natural interaction between the robots and people. The basis for appropriate human-robot interaction is that the robot should be able to detect and recognize the persons presents in its vicinity. In domestic environments, persons might be in a variety of different situations and circumstances. In this case, a robust and non-intrusive person recognition method is one of the key components of human-robot interaction.

Face is one of the most important biometric for person recognition that has widely been studied in the past. Some of the researches have been reviewed in the literature [4, 5]. However face is vulnerable to pose, illumination variations and low resolution. Furthermore, in some cases the face of a person might be occluded or not visible. For example, it is not possible to recognize and detect a person from his/her back when his face is not visible.

The other cues which make it possible to distinguish between people are global shape of the body and skeleton data known as a soft biometric features [6]. These soft biometrics can be used when other cues are not available or not enough for obtaining accurate results in long-time span. In recent years, cheap depth sensing devices such as the Microsoft Kinect have become available and the majority of modern mobile robots are equipped with them. With the appearance of these sensors, obtaining the skeleton data such as height and limb length in mobile cameras for recognition become possible. The recognition based on body information has some drawbacks. It is not completely distinctive and in some cases the person's body may not be visible or the determined skeleton data may not be reliable.

The research by Robbins et al. on humans [7] shows that people when the face is not visible used the other information such as the body information for recognizing the other people. Furthermore the recognition rate between people can be improved when both of body and face information are available and using these information is the best condition for recognition among the people. Similar to humans we used both of the body and face cues for person recognition by robot. As mentioned before, using the body and face information alone has some drawbacks and restrictions. The combination of face and body information can overcome these restriction. This information can be utilized as a non-invasive biometric for automatic recognition in various poses.

When the person's face is visible it can be used as a highly discriminative cue for recognition. While, if illumination changes and occlusion occur, the performance will be reduced. On the other hand, the information of the individual's body is less affected by illumination changes and it can be extracted in a wide variety of poses and situations. In addition, in some cases, the skeleton features may have less reliability but the combination of body information with face were shown to lead to high performance and reliability in recognition.

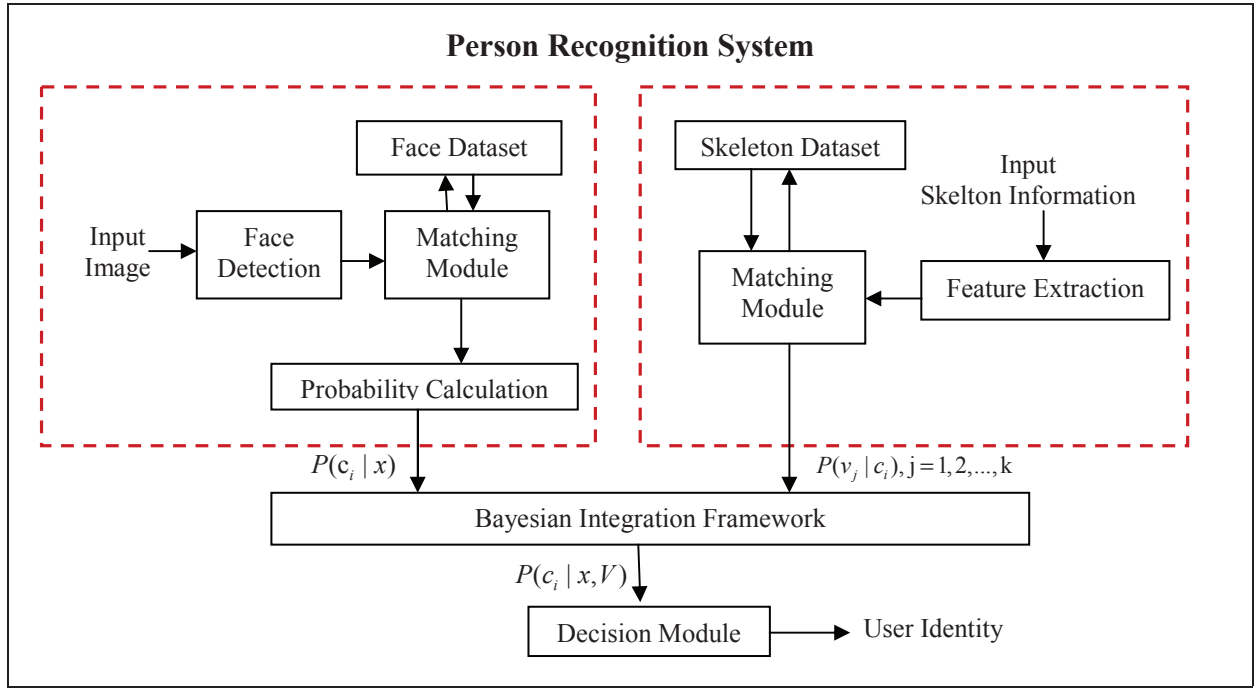


Fig. 1. Framework of integration of face and body recognition system to improve the accuracy of recognition system

This paper reports the results of a human recognition system for domestic robots based on the individual's face and body information. The main contribution of this work is combination of face and body information by using a probabilistic model for person recognition in different poses in real-time.

This paper is organized as follows: the next section reviews state-of-the-art work which is related to our approach. The proposed person recognition system is explained in section III. Experimental setup and results are presented in section IV. Finally conclusions are drawn and future work proposed in Section V.

II. REALATED WORK

Face is a promising biometric for use in person recognition. A comparative study of face recognition methods suitable for unconstrained environments is presented in [8]. Face has been used in most of the previous studies on person recognition by robots. In the following, we review some of these methods. In [9] a face recognition method based on hierarchical graph matching (HGM) has been proposed for use in mobile robots. It includes a set of nodes and connecting edges and represents the facial geometry. During the matching process, the graph tries to locate itself onto the face so that its nodes converge to predefined landmarks.

In another research by Mauricio Correa et al. [10] histograms of local binary pattern (LBP) features have been used for face recognition in thermal and visible-spectrum images. The recognition is performed using a nearest neighbor classifier in the computed feature space. Tian et al. [11] boosted the eigenfaces approach using cloud computing for face recognition by robots.

Scale invariant feature transform (SIFT) features are used in [12] for face recognition. These features are extracted only from selected regions of the face and afterwards the matching

process is performed only between the corresponding regions. In [13] the author endows the ability of recognizing people by unsupervised face recognition based on eigenfaces method in a humanoid robot.

Similar to an individual's face information, the body information might be used for recognition. There are only a few works that focus on recognition based on the depth information for extracting skeleton features because it is not a long time since cheap 3D sensors have become available. In a research by Meifen Cao et al. [14] PCA is used to extract an individual's body features for recognizing him. This method applies PCA to 19 link lengths of the individual. Then the individual's feature vectors are obtained and used for recognition and tracking of the individual, by the robot.

Transformation of the whole point-cloud of the body to a standard pose based on the skeleton tracker information is proposed in [15]. In this method, using Iterative Closest Point (ICP), the body shapes of individuals are matched for re-identification. In [16] skeleton-based and surface-based features are utilized for recognition. In this approach, a vector of features from geodesic distances on the body shape and skeletal data as the lengths of a few limbs of a person is extracted. Persons are classified using nearest neighbor method.

The use of head and shoulders area in 3D range data for person recognition is proposed in [17]. In this approach, in situ person recognition has been performed based on a support vector machine (SVM) in which the head to shoulder signature (HSS) is utilized as the basis. Person recognition by a humanoid robot is performed by combining face and speaker recognition methods in [18]. These two methods are combined together in score and decision level with equal weights for each of these models.

III. THE PROPOSED SYSTEM OVERVIEW

The proposed recognition system includes two subsystems: the primary one which is based on face and the secondary one which is based on the body information (Fig. 1). The system consists of three main components: 1) face detection and recognition, 2) feature extraction from the skeleton data, and 3) combining the results of each model of the skeleton and face with a probabilistic method.

A. Face Detection and Recognition

For face detection we used the Viola-Jones method [19]. This method is the most popular face detection paradigm which is based on the use of cascades of boosted classifiers, which allows robust and efficient detection of faces. By using Viola-Jones algorithm, the output would be a set of detected faces in the image.

We used the sparse representation technique for face recognition proposed by Wright et al. [20]. This method is selected because of its high performance reported in former studies, in addition to being real-time. The evaluation considered real-world conditions that include facial expressions, occlusions, and illumination changes. The input face image is reconstructed for each class based on the sparse coding coefficients of the class.

In this approach, the training samples of individual faces are considered as the overcomplete dictionary. Let $D = [X_1, \dots, X_c, \dots, X_C] \in \mathbb{R}^{d \times N}$ be the set of original training samples, where C is the number of classes of individual faces and $X_c \in \mathbb{R}^{d \times N_c}$ is the sub-set of all the N_c vector-represented training samples from class c . Then the input test face is represented as a linear combination of the training samples. The combination is sparse in comparison to the whole training samples. Let $x \in \mathbb{R}^d$ be a query facial image, then x is sparsely coded over X via l^1 -norm minimization as:

$$\hat{a}_i = \arg \min \|a\|, \quad \text{subject to } Da = x. \quad (1)$$

where, a is the sparse coefficient vector. For classification the minimum reconstruction error of the input testing image with the sparse coding coefficient is evaluated. Hence, the minimum reconstruction error is calculated by:

$$e_i = \arg \min_i \|x - X_i \delta_i(\hat{a})\|_2, \quad (2)$$

where, $\delta_i(\cdot)$ is a vector indicator function that extracts the elements corresponding to the i^{th} class. The test image is classified into the training classes with minimum reconstruction error.

Our approach operates under the assumption that the domestic robots interact with only known people at home. The probability of the degree test image belongs to each of the training class is calculated using the reconstruction error evaluation. Thus the output of the primary biometric system is of the form $P(c_i | x)$, $i=1,2,\dots,n$, where $P(c_i | x)$ is the probability that the test image is c_i given the feature vector x . This probability is calculated based on the maximum $e_{\max}$ and the minimum $e_{\min}$ reconstruction error obtained from

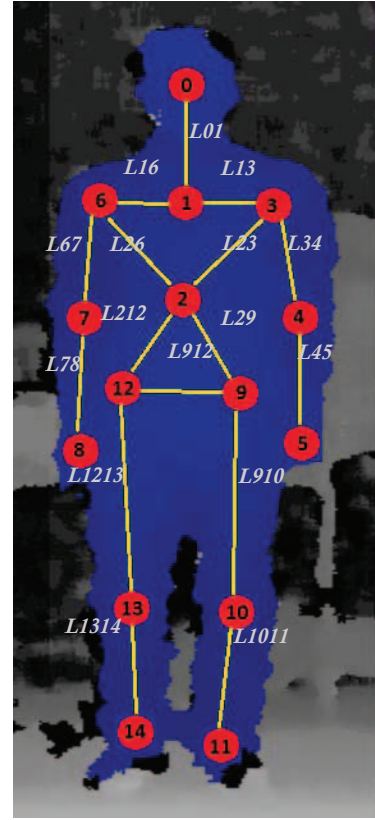


Fig. 2. Link length between adjacent joints

comparing test image with each of the training classes. Based on these values the probability of each classes is computed as:

$$P(c_i | x) = \frac{e_{\max} + e_{\min} - e_i}{n(e_{\max} + e_{\min}) - es} \quad (3)$$

where, n is the number of training classes and es is the summation of error values of test image with all training classes.

B. Feature Extraction From Skeleton Data

In this research, the skeletal tracker developed by OpenNI [21] is used to obtain coordinate of 3D positions for 15 different skeletal joints in real-time. The position of 15 joints obtained from the OpenNI tracker is shown in Fig. 2. This tracker works on 3D data and has a superior performance over the 2D-based approaches.

In this paper recognition is not only done in the standing still state, but also in the moving and sitting state of a human. Thus coordinate of the 3D skeleton position is acquired during all states. The skeleton data acquired in one measurement which includes standing still and each of the other states is called one sample sequence in this paper.

In this section, features are extracted from the 3D position coordinate of 15 skeletal joints. Our skeleton feature vector consists of 14 elements, each of which is composed of the link length between some adjacent joints and the distance summation of some joints. The link lengths between adjacent joints are shown in Fig. 2, in which the distances between the joints are calculated using the Euclidean distance.

These lengths which are used for skeleton features extraction are shown in Fig. 2. The individual feature vector V_i is defined as:

$$V_i = [v_{i1}, v_{i2}, \dots, v_{i14}], \quad (5)$$

where, i stands for the i^{th} individual which should be recognized. These features are calculated as:

$$\begin{aligned} v_1 &= L_{01} & v_8 &= L_{1213} \\ v_2 &= L_{26} & v_9 &= L_{34} + L_{45} \\ v_3 &= L_{23} & v_{10} &= L_{67} + L_{78} \\ v_4 &= L_{12} & v_{11} &= L_{1213} + L_{1314} \\ v_5 &= L_{16} & v_{12} &= L_{910} + L_{1011} \\ v_6 &= L_{13} & v_{13} &= L_{12} + L_{29} + L_{910} + L_{1011} \\ v_7 &= L_{910} & v_{14} &= L_{01} + L_{12} + L_{29} + L_{910} + L_{1011} \end{aligned}$$

Standing still sequence is used to calculate the average and variance of each of these features for every individual. These values are used for calculating the probability of each individual.

C. Combination of face and body recognition

Based on the combination method proposed by [6], we combined the results of the face and body recognition systems. Let $c_1, c_2, \dots, c_n$ represent the n users who are enrolled in the dataset and $V = [v_1, v_2, \dots, v_k]$ be the soft biometric feature vector. Given the primary biometric feature vector x and soft biometric feature vector V , the probability of user c_i , i.e., $P(c_i | x, V)$ can be computed. By using Bayes rule and considering that x and V are independent from each other, the probability is calculated as:

$$P(c_i | x, V) = \frac{P(V | c_i) P(c_i | x)}{\sum_{i=1}^n P(V | c_i) P(c_i | x)} \quad (6)$$

If it is assumed that the extracted features from the body are independent, (6) can be rewritten as:

$$P(c_i | x, V) = \frac{P(v_1 | c_i) \dots P(v_k | c_i) P(c_i | x)}{\sum_{i=1}^n P(v_1 | c_i) \dots P(v_i | c_i) \dots P(c_i | x)} \quad (7)$$

where, $P(v_j | c_i), j=1, 2, \dots, k$ represents the conditional probability of the feature v_j given the user c_i . By considering Gaussian distribution for skeleton features, the $P(v_j | c_i)$ in (7) is calculated as in (8).

$$P(v_k | c_i) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(v_k - \mu_i)^2}{2\sigma_i^2}} \quad (8)$$

where, μ_i and σ_i are calculated from training data of the i^{th} class and k is the number of skeleton features. The logarithm of (7) can be expressed as:

$$\begin{aligned} \log P(c_i | x, V) &= \log P(v_1 | c_i) + \dots + \log P(v_k | c_i) \\ &+ \log P(c_i | x) - \log P(v), \end{aligned} \quad (9)$$

where, $P(v) = \sum_{i=1}^n P(v_1 | c_i) \dots P(v_k | c_i) P(c_i | x)$.

Equal weighting of the features is a major drawback of this equation. Because of this, the soft characteristics have an equal say in decision as the primary biometric trait. But in practice the face of a person has more distinctive ability and performance for recognition compared with the skeleton features. Consider a situation that the face recognition is done correctly, but a specific soft biometric trait of a user is measured incorrectly. In this case, a correctly recognized person is highly possible to be rejected only because of equal weighting of the features. Therefore we need a weighting scheme to combine these biometrics. Using (10), the proposed weighting scheme for user c_i is:

$$g_i(x, V) = b_0 \log P(c_i | x) + b_1 \log P(v_1 | c_i) + \dots + b_m \log P(v_m | c_i), \quad (10)$$

where, $\sum_{j=0}^m b_j = 1, b_0 \gg b_j, j=1, 2, \dots, m$. The assigned weight for face is b_0 and b_j 's, $j=1, \dots, m$ are the weights assigned to the features that are extracted from the body.

Considering this equation, only if several soft biometric traits do not match will result significant reduction in the posterior probability and the user specified with the face recognition system could be possibly rejected.

When one of the sub-components is reporting there is no data (face or any of the skeleton features), with which to make a decision and calculate the probability the combined method defaults to the values of the remaining component. For example, if the face of the person is not visible to the robot's camera, then the skeleton features recognition results used alone.

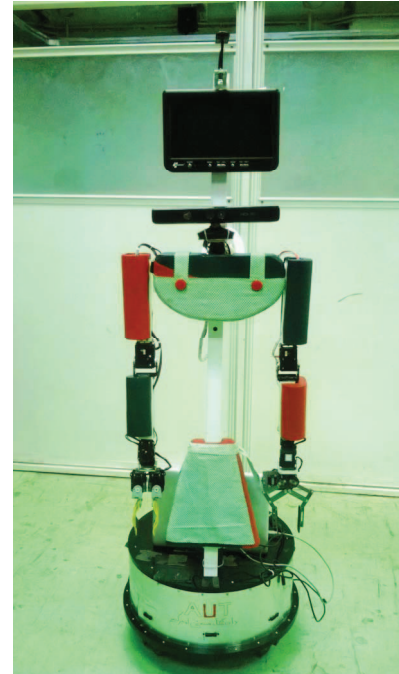


Fig. 3. Sepanta, the service robot of Amirkabir University of Technology. This robot is equipped with a Kinect for providing 3D data.

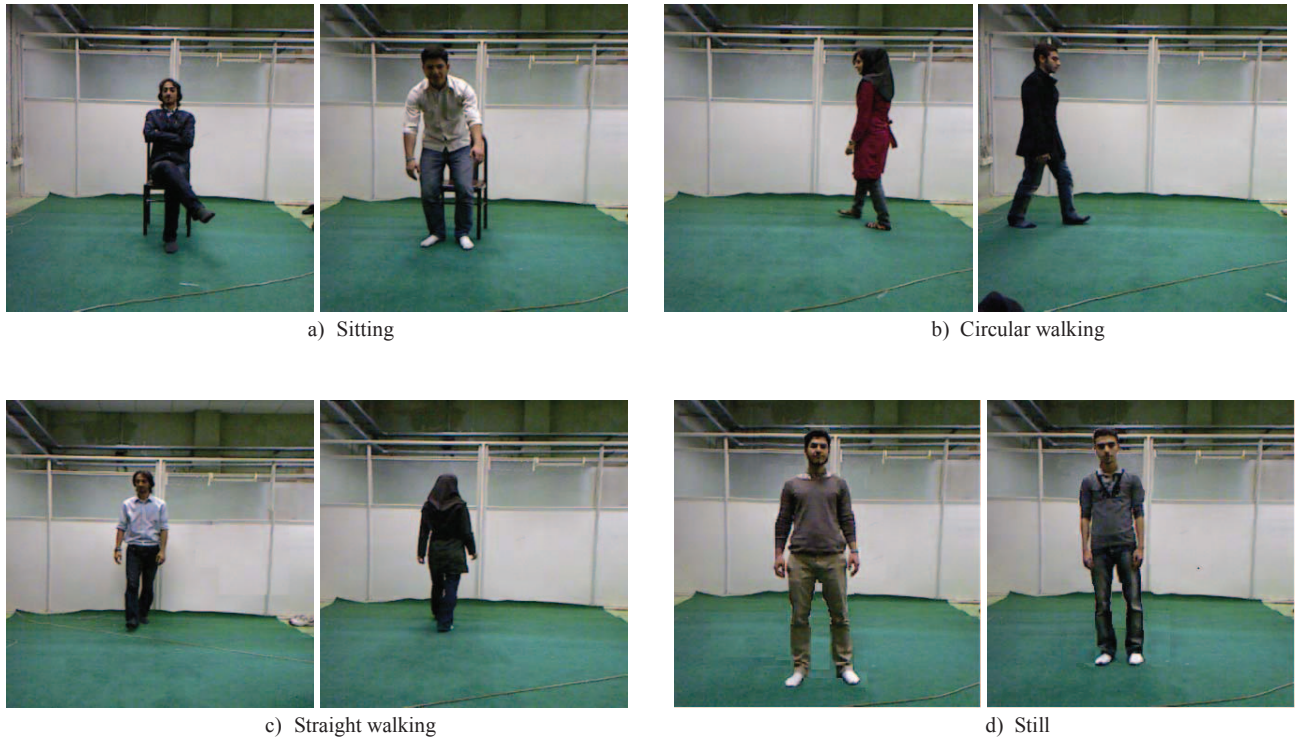


Fig. 4 Sample images of persons in four different poses that were used for training and recognition by Sepanta

IV. EXPERIMENTS

This section presents results from an empirical evaluation of the person recognition system in different situations based on the body and face information. The implementation has been packaged as a node in the open-source Robot Operating System (ROS) [22], allowing it to be used on any robot that is equipped with Kinect. In the following section, our platform for evaluating the algorithm is first introduced, and then the results are given.

A. Robot Platform

To evaluate the proposed algorithm we have used a service robot named “Sepanta” (Fig. 3). This robot is equipped with a Kinect sensor which provides 3D visual information, including an RGB image sequence at a frame rate of 30HZ, which enables Sepanta to detect and recognize the person’s face, and depth information which produces the 3D position of the joints of the person.

B. Experimental Results

There are usually a few individuals in domestic environments. Hence the proposed method is evaluated only on a five person group; three adult men and two adult women. In this experiment, we focused on recognizing persons while they are in different poses so that in every pose, only some of the body information or face is available. The data for evaluation was collected in different days. The previous methods used each of face or body cues lonely to recognize persons by the robots. Thus we classified the experiments into three categories: firstly, we evaluated the performance of the face

system alone; secondly the body recognition system alone, and thirdly, the combined system.

We consider four different sequences of straight walking, circular walking, sitting, and standing still. Standing still is used for training and the other three sequences are used for test. First, in the sitting sequence both the body and face information are available, but the skeleton information cannot be determined with high accuracy. In addition the face resolution is low for recognition. Secondly, in walking straight, when the person walks towards the robot, both the face and the body are visible, but when the person walks away, the face is not visible and only the body information is accessible. Thirdly, when the person walks in a circular path, some joints of the body are not visible and the face is visible from the side view. Some of example images of persons in different poses are shown in Fig. 4.

In the still pose, each person stands in front of the camera for five seconds and the data are saved into a dataset to be compared with the test sample. For each person, 5×30 frames will be available. From every 15 frames, an image of the face is saved, resulting in ten images of the face for every person. The data for each test sample sequence is collected from the moment the persons appear in the scene until 20 seconds later. For each person, the skeleton features are computed. Then by taking the average of them, a vector of 14 skeleton features is formed and saved into the dataset for each person.

The discriminative features is different for each family for example in a family the height of members can have a high

TABLE I. RECOGNITION RATES OF FACE AND BODY RECOGNITION SYSTEM AND THEIR COMBINATION

	Sitting(%)	Circular walking(%)	Walking straight(%)	Average(%)
Face	76.84	10.19	34.42	40.48
Body	44.93	71.61	85.15	67.23
Face-body	81.69	72.33	86.69	80.23

discriminative ability but in the other one the length between shoulder and torso. Thus it is necessary to adapt the importance of each features based on the family members. For this reason to obtain b_j 's (the weights that was described in sec III) the training data was separate in two parts, the first part was used for the system training and the second part used to get the recognition rate for each feature individually. Considering these recognition rates the weight for each feature is determined.

Table I shows the person recognition rates for the face system, body system, and the combination of them. It can be seen that the combination of the face and skeleton system has produced significant improvement in performance in comparison to the face system and body system in the sitting state. Since the mentioned systems are independent, using them simultaneously overcomes the weak points of both of them. In the circular and straight walking the face is partially visible so the face recognition rate will decrease and subsequently the face system does not compensate false results produced with body system. By taking the average of all the sample sequences, we get the accuracy of 40.48% for the face recognition system, 67.23% for the body system and 80.23% for the combination of them. The results show that using the face and the body systems together will lead to a higher performance.

V. CONCLUSION

This paper presents a method for person recognition in different poses for a domestic robot in real-time. The body features are utilized and combined with the face information for solving the problems of person recognition. Integration of face and body biometrics is done using the Bayesian decision theory. It is shown that the proposed method can recognize persons in cases that the face or body information is not available. Also we demonstrate that the combination of these biometrics traits leads to a higher accuracy in establishing the user identity. This method is applied on a service robot by using its RGBD sensor (Microsot Kinect).

In future work, we will study recognition of partly occluded bodies and in more varied poses. To improve the recognition accuracy in various situations, we also plan to develop an algorithm to choose the features and adapt the weight of features based on the context.

REFERENCES

- [1] J. Sung, R. E. Grinter, and H. I. Christensen, "Domestic robot ecology," in *International Journal of Social Robotics*, Vol. 2, No. 4, pp. 417-429, Jul. 2010.
- [2] Breuer, Thomas, Geovanny R. Giorgana Macedo, Ronny Hartanto, Nico Hochgeschwender, Dirk Holz, Frederik Hegger, Zha Jin et al. "Johnny: An autonomous service robot for domestic environments," in *Journal of Intelligent & Robotic Systems (JIRS)*, vol. 66, No. 1-2 pp. 245-272, 2012.
- [3] C. Ray, F. Mondada, and R. Siegwart, "What do people expect from robots?," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3816-3821, September 2008.
- [4] J. H. Shah, M. Sharif, M. Raza, and A. Azeem, "A survey: Linear and nonlinear pca based face recognition techniques," in *International Arab Journal of Information Technology*, Vol. 10, No. 6, pp. 536-545, 2013.
- [5] T. Sinsinwar, and P. K. Dwivedi, "A survey on feature extraction method and dimension reduction techniques in face recognition," in *International Journal of Applied Science and Engineering Research*, Vol. 3, No. 3, pp. 734-740, 2014.
- [6] A. K. Jain, S. C. Dass, and K. Nandakumar, "Soft biometric traits for personal recognition systems," in *Biometric Authentication*, pp. 731-738. Springer Berlin Heidelberg, 2004.
- [7] R. A. Robbins, and M. Coltheart, "The effects of inversion and familiarity on face versus body cues to person recognition," in *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 38, No. 5, pp. 1098-1104, 2012.
- [8] J. Ruiz-del-Solar, R. Verschae, and M. Correa, "Recognition of faces in unconstrained environments: a comparative study," in *EURASIP Journal on Advances in Signal Processing*, Vol. 2009, pp. 1-19, 2009.
- [9] U. Handmann, S. Hommel, M. Brauckmann, & M. Dose, "Face Detection and Person Identification on Mobile Platforms," in *Towards Service Robots for Everyday Environments*, pp. 227-234. Springer Berlin Heidelberg, 2012.
- [10] M. Correa, G. Hermosilla, R. Verschae, and J. Ruiz-del-Solar, "Human detection and identification by robots using thermal and visual information in domestic environments," in *Journal of Intelligent & Robotic Systems*, Vol. 66, No. 1-2, pp. 223-243, 2012.
- [11] S. Tian, D. Saitov, and S. G. Lee, "Cloud Robot with Real-Time Face Recognition Ability," in *Advanced Science and Technolog Letters*, Vol. 51, pp. 77-80, 2014.
- [12] C. Cruz, L. E. Sucar, and E. F. Morales, "Real-time face recognition for human-robot interaction," in *8th IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 1-6, September 2008.
- [13] L. Aryananda, "Recognizing and remembering individuals: Online and unsupervised face recognition for humanoid robot," in *IEEE/RSJ International Conference on, Intelligent Robots and Systems*, vol. 2, pp. 1202-1207, 2002.
- [14] M. Cao, and H. Hashimoto, "Specific person recognition and tracking of mobile robot with Kinect 3D sensor," in *39th IEEE Annual Conference on the Industrial Electronics Society (IECON)*, pp. 8323-8328, November 2013.
- [15] M. Munaro, A. Basso, A. Fossati, L. Van Gool, and E. Menegatti, "3D reconstruction of freely moving persons for re-identification with a depth sensor," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4512-4519, 2014.
- [16] I. B. Barbosa, M. Cristani, A. Del Bue, L. Bazzani, and V. Murino, "Re-identification with rgb-d sensors," in *Computer Vision-ECCV Workshops and Demonstrations*, pp. 433-442. Springer Berlin Heidelberg, 2012.
- [17] N. Kirchner, A. Alempijevic, and A. Virgona, "Head-to-shoulder signature for person recognition," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1226-1231, 2012.
- [18] E. Martinson, W. Lawson, and J. G. Trafton, "Person identification through perceptual fusion," in *10th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, pp. 358-364, December 2010.
- [19] P. Viola, and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 1-511, 2001.
- [20] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 2, pp. 210-227, 2009.
- [21] OpenNI, OpenNI. <http://openni.org/> (Last Accessed: February 2015).
- [22] M. Quigley, K. Conley, B. P. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "Ros: an open-source robot operating system," in *ICRA Wshop Open Source Soft.*, 2009.