Real-Time Recognition of Human Postures for Human-Robot Interaction

Zuhair Zafar, Rahul Venugopal*, Karsten Berns

Robotics Research Lab
Department of Computer Science
Technical University of Kaiserslautern
Kaiserslautern, Germany
Email: {zafar,berns}@cs.uni-kl.de, venugo@rhrk.uni-kl.de*

***Abstract*—To function in a complex and unpredictable physical and social environment, robots have to apply their intellectual resources to understand the scene in an efficient and intelligent way, similar to humans. Especially when interacting with humans, this cognitive task becomes more challenging. The work in this paper is focused on recognizing human actions and postures during daily life routines in real-time to understand human motives and emotions during a dialogue scenario. Using depth data, a real-time approach has been proposed that uses human skeleton joint angles to recognize 19 different human postures (standing and sitting). Feature vectors are constructed after pre-processing of joint angles. A supervised learning mechanism has been used to train the classifier using Support Vector Machine. Approximately 30000 training samples have been created for training purpose. The system recognizes all the postures accurately provided the skeleton tracker is working precisely when tested on the database. During live testing, the system reports 98.2% recognition rate, proving the potential of the proposed approach.**

Keywords–Human-robot interaction; skeleton data; human posture recognition; feature vector; classification.

I. Introduction

Human posture recognition is an active research topic in the field of human-robot interaction. In addition to being used in the context of humanoid robotics, recognition of human postures has many applications in human assistive systems and in the automobile industry. The basic objective is to enable humanoid robots to work side by side with humans during daily life. In order to realize this goal, robotic systems must have the capability to differentiate human(s) from the cluttered environments. In addition to detecting humans, these systems should also analyze their posture, actions, emotions, motives and overall behavior. This, in turn, helps the robots to be more intelligent and resourceful when interacting with humans. Human behavior can be analyzed using human's nonverbal communication. According to [1], two thirds of our communication consists of non-verbal communication and only one third of our communication consists of verbal content. Nonverbal communication consists of facial expressions and bodily cues. Human posture represents an important part of the nonverbal communication.

Human posture and body movement play significant role in the perception of interaction partner. Humans use different hand gestures and body postures to express their internal emotional state in different situations. In humans, postures provide significant information through nonverbal cues. Psychological studies have also demonstrated the effects of body posture on emotions. This research can be traced back to Charles Darwin's studies of emotion and movement in humans and animals [2]. A massive study and research has been conducted in 1970s on the significance of body language in which the main area of focus was leg-crossing, defensive posture and arm-crossing, suggesting that these nonverbal behaviors depict feelings and attitude. Posture can also rely on the situation, i.e., people change their postures depending on the situation. Currently, many studies have shown that certain patterns of body movements are indicative of specific emotions [3][4]. Researchers have studied sign language and found that even non-sign language users can determine emotions from only hand and body movements [5]. For example, anger is characterized by forward body movement [6].

Posture recognition plays an important role in expressing human emotions. Many scientists believe that all the variations of postures are due to the change in emotions and play significant role in human evolution. Human emotions are always difficult to understand and there are many factors that influence human emotions. The art of recognizing human emotions had gained its importance long back and is, currently, studied actively [7]. Some behavioral cues can be easily recognized from postures. For example, a person scratching his head during interaction shows thinking behavior. Similarly, crossed arms posture shows that the interlocutor is reserved and is trying to block himself from opening to other person.

However, the challenge is to recognize complex human postures in cluttered environment in real-time, especially, those set of postures which are used in daily life in human-human interaction scenarios. For example, crossed arm, pointing with left or right arm, casual or attentive standing posture, relaxing posture, thinking or shrug posture, etc. On the contrary, every region or culture has its own different postures which, sometimes, are totally opposite in meaning in some another culture. One of the major challenges in recognizing human postures is diversity in people performing postures. People from different culture are expressing the same posture in different ways as compared to others. In addition, postures are also dependent on the height and human physique variations which make them more challenging to recognize. Moreover, sitting postures appear different from standing postures and need separate classifier for the posture recognition task.

Numerous ways have been reported in the literature to recognize human postures. Some of these methods use wearable sensors to extract the psychological parameters like electroencephalography (EEG) data, skin temperature, accelerometer readings, etc. However, these methods require special sensors to wear all the time and sometimes require training how to use them. In contrast, approaches using visual information from the visual sensors are more natural means of recognition of human posture. However, this work explores recognition of human postures using RGB and depth (RGB-D) sensor. This work uses ASUS Xtion [8], installed on a humanoid robot, ROBIN [9] to extract distance data. With the help of OpenNI and NiTE library, the system is able to extract human skeleton joints. These joints are then pre-processed and converted into angles to make the system invariable to human height or physique. Feature vectors are generated using angle information between each joint and classified using Support Vector Machines (SVM). The major contribution of the paper is the accurate and automatic recognition of human postures in real-time using kinect-like sensor. Our approach reports close to 100% results when the human skeleton is tracked accurately in real-time in cluttered environment. Moreover, the system is also capable of distinguishing between standing and sitting human postures using human height analysis. In the following sections, we describe the overall approach and experimental results in detail.

The rest of the paper is organized as follows: related work is discussed in Section 2, Section 3 discusses human posture recognition approach and classification in detail. Experimental results and performance evaluation are discussed in Section 4. We conclude the paper in Section 5.

II. Related Work

Research on posture recognition using skeleton data began in the 1990s and is still being carried on. Generally, posture recognition approaches can be separated in two broad categories: (a) wearable sensors based posture recognition and (b) posture recognition using vision based sensors. Wearable sensors include gloves and other commercially available products that are used to extract different statistical and geometrical information of the limbs or body when worn. Few of these devices namely Sensewear, ActiGraph and ActivPal have been used by Wang et al. [10]. They address challenges like data imbalancement, instant recognition and sensor deployment in order to achieve an overall accuracy of 91% for sitting, standing and walking postures. Similar approaches using wearable sensors have been reported with higher accuracy. However, these require sensors to be worn. Latter approaches use vision sensors for the recognition of human postures. The advantage of this approach is twofold: first, these approaches are noninvasive; and secondly, they are also cost-efficient. Humans can perform their gestures and postures in front of a camera sensor without any other device attached to their bodies for posture recognition tasks.

Posture recognition via vision sensors can be further divided into two categories namely camera based posture recognition and RGB-D sensor based posture recognition. Numerous works have been reported in the literature that use monocular camera to estimate human pose and human action. The most general approach is to extract features from images based on the structure of the human body, e.g., skin color or face position [11]. However, this approach impose restrictions on features such as clothes and orientation. There are other methods to extract silhouettes and edges as features from the image [12][13]. However, they rely on the stable extraction of the silhouettes and edges. Moreover, they perform poorly in self-occlusion.

In order to address these shortcomings, researchers use depth sensors to extract human joint positions. S. Nirjon et al. [14] describe a system, called Kintense, which is real-time system with a high accuracy to detect human aggressive actions, e.g., hitting and pushing that are relevant for games. The system has been trained using supervised and unsupervised machine learning techniques. The sensors calculate distance between body and the cameras, skeleton joints and speed at which an action is performed. Deep learning and neural networks are used to eliminate false positives and to identify actions that are not labeled. Real-time testing has been performed by deploying the system in more than one multiple-person household which illustrates the sensitivity of the system towards unknown and unseen actions. The real-time system proves that the accuracy of the system is more than 90% [14].

Using RGB-D sensor, Zhang et al. [15] extract joint positions of a human with the help of Microsoft Kinect. In order to make it independent of human size, each joint position is normalized using its neighboring joint to make a feature. This feature vector which consists of all normalized joints is then classified using SVM. A total of 22 postures are recognized with 3 different classifiers. The drawback of this approach lies in normalization of joint positions. Although authors claim that the system is invariant to human size, it would not be invariant to human height or size of the limbs completely as normalization only adjusts joint values with its neighboring joint.

Another similar work has been conducted by Ivan Lilloa et al. [16] to recognize human activities using body poses estimated from RGB-D data. The system modules are classified into three different levels which include geometry and motion descriptors at the lowest level, sparse compositions of these body movement at the intermediate level, spatial and time stamped compositions used to represent human actions involving multiple activities at the highest level. The work is related to dictionary learning method and their framework focuses on vector quantization using k-means to cluster low-level key point descriptors for dictionary learning [17]. The model developed uses an alternative quantization methods, discriminative dictionaries, or different pooling schemes [18]. Sparse coding methods have also been used for alternative quantization methods. These methods have mostly focused on non-hierarchical cases where mid-level dictionaries and top-level classifiers are trained independently [17]. Niebles et al. [18] extend this model to the case of action recognition. In contrast to former approach, the model is limited to binary classification problems and reports good accuracy only in a constraint scenario.

In previous related work, the required data is captured either from images or videos and the processing is done to create the feature vector. Feature vector represents the data in a form such that the system can be trained. Many classification techniques have been used in classification of the training dataset, such as SVM, neural networks and deep learning techniques. After the classification, the system can be tested offline using existing database or online testing in real-time scenario. Most of these approaches are used only to recognize standing postures or actions. Additionally, these approaches are not robust to real-time recognition of human postures with

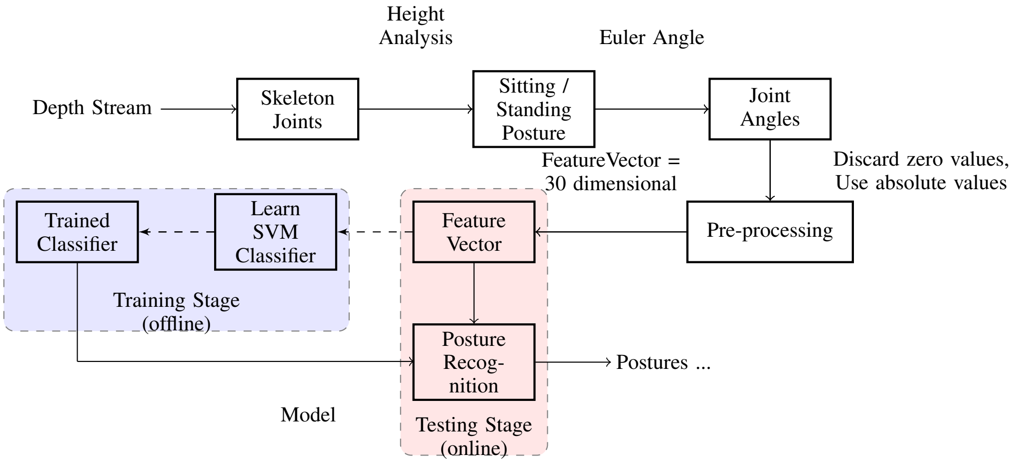

Figure 1. Working schematics of the approach. Using depth stream and NiTE Library, skeleton joints are detected. Based on the height, system classifies the subject either standing or sitting, after which joint angles are computed from joint positions to construct a 30 dimensional feature vector for classification.

more than 10 classes. In this paper, we have proposed an approach that is robust to real-time recognition of postures and can differentiate between standing and sitting postures. Moreover, it can recognize 19 postures used in daily life routine. The detailed analysis of the proposed approach is discussed in following sections.

III. Human Posture Recognition

Visual perception in complex and dynamical scenes with cluttered background is a challenging task which humans can solve remarkably well. However, it performs poorly in this kind of challenging scenarios for a robot perception system. One of the reasons of this large difference in performance is the use of context or contextual information by humans. Furthermore, robot has to perform its computations as fast as possible due to the notion of real-time. As a result, most of the time robot perception system is hampered with low resolution images. There is a need to develop such perception system which can cater complex environments and work efficiently.

This paper presents an approach that uses depth data along with NiTE library to detect human joint positions and then convert them into meaningful angles for feature vector generation task. The resultant feature vector is quite unique for each posture and is invariant to height, body shape, illumination, proximity and appearance of human. The working schematics of the proposed approach is presented in Figure 1. Our proposed approach reports high accuracy for both sitting and standing postures. The system is able to recognize overall 19 gestures real-time when classified by using multi-class SVM. Each module of the approach is described in the following sub-sections.

A. Depth Image

Instead of using monocular camera, ASUS Xtion is employed in order to utilize depth data. The advantage of using such devices with depth sensor lies in the segmentation of human skeleton using OpenNI and NiTE Library. Segmenting humans on the basis of silhouette and edges might work in a constraint scenario but it behaves poorly when applied in dynamic environment. In contrast, human can be detected and tracked efficiently using depth sensor in constantly changing scenario with a lot of different daily life objects involved. This sensor can work efficiently in the range of 0.5 to 3.5 meter.

B. Skeleton Data and Joint Positions

Fifteen different skeletal joint positions of human can be extracted in real-time using OpenNI and NiTE libraries. These joint positions are quite accurate and tracked over time. Furthermore, NiTE middle ware library allows multiple human tracking and joint positions extraction in real-time. In order to extract joint positions reliably, the whole human body should be clearly visible to RGB-D sensor with no complete occlusions of body parts. The disadvantage in using joint positions is the dependence on correct detection of human skeleton. Due to partial occlusions of limbs, the module can report ambiguous skeletal information which effects the joint position values. Figure 2 shows tracked humans with their respective skeletal joints.

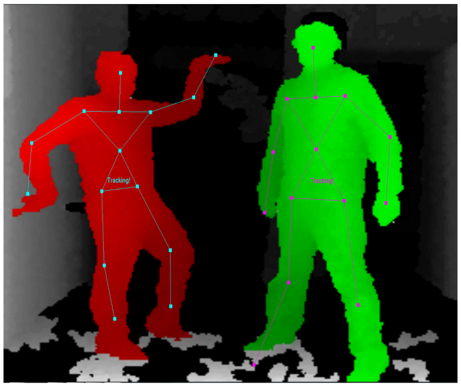

Figure 2. Multiple tracked humans and their skeletal joints. (Image used from www.openni.ru/files/nite/index.html)

C. Sitting or Standing Postures

Before recognizing postures, the important step is to detect whether human is standing or sitting. The simplest way is to analyze the height of human with respect to its z distance from the sensor. Empirical studies have shown that the relation between these two entities is linear. For example, if human is near to the sensor, he/she appears taller and similarly, if human is away from the sensor, he/she appears short. To make it height and scale invariant, the system uses the depth data (z distance) to normalize the height of the person. If the human head joint has the value more than the set threshold value, system would classify it as standing posture. If he/she has the head joint position value less than the set threshold, the system would classify it as sitting.

D. Joint Angles

The major disadvantage in using joint positions for feature extraction task is that they are variant to positions, height and limbs variations. This type of features might report better results when the position and height of the human would be fixed. However, these features behave poorly when dealing with varied height or dynamic humans. In order to solve this problem, researchers have proposed an approach that calculates distance of each joint from torso to make a feature vector. Although this type of feature extraction reports better results, it is still dependent on the height of the person.

In order to address this shortcoming, this paper proposed a unique method to extract features. Instead of using joint position for feature extraction task, these joints positions are converted into angles between each two joints. The benefit of using angles is that they are not dependent on the position or height or human physique, instead they compute directions between each joint. The direction between each joint would be similar for a short person and a tall person if they are expressing the same posture. Euler angles are used to convert the joint positions to angles. Following (1) - (3) are used to compute angle between joint a and b.

$$angle_x = \tan^{-1} \frac{(a_y - b_y)}{(a_x - b_x)} \tag{1}$$

$$angle_y = \tan^{-1} \frac{(a_z - b_z)}{(a_y - b_y)} \tag{2}$$

$$angle_z = \tan^{-1} \frac{(a_x - b_x)}{(a_z - b_z)} \tag{3}$$

The angles are then converted from Radian to degrees using (4).

$$angle_x^{\circ} = angle_x * 180/\pi \tag{4}$$

E. Pre-processing and Feature Extraction

In total, 15 joint angles can be calculated for each posture. However, it has been observed that certain joints do not contribute in deciding the posture. Joint angles between knee and foot, or hip and knee do not add useful information for posture recognition task. The reason lies in the postures, recognized in this work, are not effected by joint angles of lower body. During this pre-processing stage, the number of angles recorded are reduced to 10.

On the other hand, NiTE library can detect and track human but it is not able to distinguish whether human is facing towards the camera or his/her back facing the camera. This makes the direction of angles totally opposite. In order to make the system invariant to human facing direction, we take the absolute value of all the joints, thus making the features more consistent for the same class. Joint angles with values 0°, 90°, 180° or 270° in 10 consecutive frames are also discarded. After empirical studies, it has been found out that when part of the limb or body is occluded, skeleton tracker reports $(0, 0, 0)$ joint position. This leads to false recognition, therefore, the instances are discarded. 10 joint angles are then used to construct a feature vector. Since every joint angle has x, y, z values, the feature vector for a single depth observation becomes 30 dimensional.

F. Classification

Classification is an important step in any recognition task. The major task of classification stage is to differentiate each class or category accurately based on the knowledge gained during the training stage. Numerous classification algorithms have been presented in machine learning, e.g., neural networks, decision trees, random forests, convolution neural network, etc. This work uses SVM, a supervised learning algorithm, for the classification task. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall [19]. The benefit of SVM lies in the regularization parameter which if set accurately, avoids overfitting. Moreover, it uses the kernel trick, i.e., it can build an expert knowledge about the problem by engineering the kernel. SVM generalizes on high dimensional feature set quite well given that the database is also huge.

This paper uses multi-class SVM classification. More than 2100 instances are used during training stage for each posture and 40000 instances for the whole training data are used for 19 classes. 10 different subjects, from different ethnicity (Indian, Pakistani, German, Italian and Turkish), featured in the training dataset. Linear kernel with regularization parameter $C = 0.4586$ is used during SVM training. Figure 3 shows 3D graphical plots of joint angles between *right_shoulder-right_elbow* and *right_elbow-right_hand*. It can be seen that the classes are easily distinguishable based on the angle between two joints. With the contribution of other joints angles between joints, the problem is easily classified by SVM linear kernel.

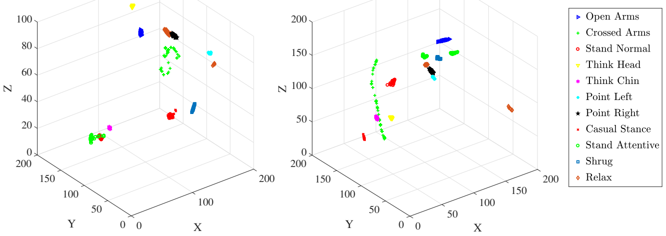

Figure 3. Samples from training data in 3D plane for each class marked with different color. (a) Angle values between right shoulder and right elbow (b) Angle values between right elbow and right hand.

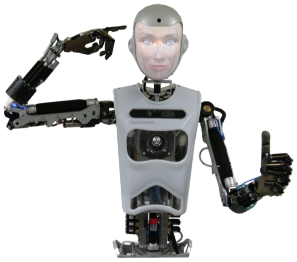

Figure 4. ROBIN - Humanoid robot of TU Kaiserslautern

IV. Experimentation and Evaluation

The goal of the system is to recognize human postures in real-time robustly in order to realize human-robot interaction. The humanoid robot, named ROBot-human-INteraction (ROBIN), is used in order to evaluate the posture recognition system. ROBIN has been developed by Technical University of Kaiserslautern [9] as shown in Figure 4. It consists of intelligent hands that can express almost any gesture. Single whole arm has 14 degrees of freedom that uses compressed air to perform any action. Head and torso of ROBIN also have 3 degrees of freedom. The backlit projected face is able to express different expressions and emotions. Additionally, ROBIN can speak in English and German using text-to-speech software. ASUS Xtion is installed on the chest of ROBIN, which is used for all the perception tasks, e.g., posture recognition, gestures recognition, etc. ROBIN has its own processor that can handle all the movements of joints. In the following subsection, a detail analysis of postures and experimentation are discussed.

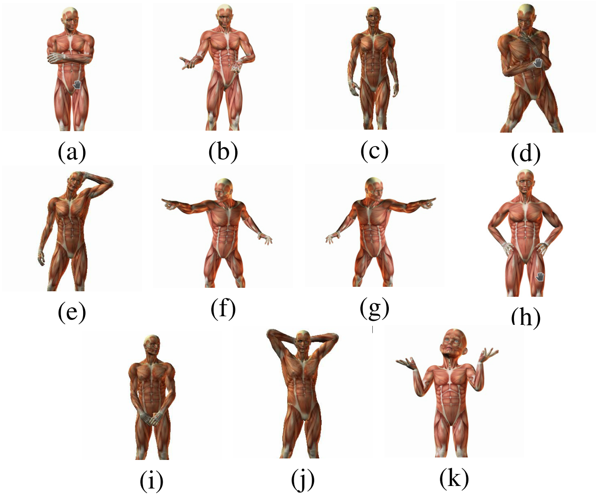

Figure 5. Standing postures (a) Crossed Arms (b) Open Arms (c) Stand Normal (d) Think (Hand on chin) (e) Think (Hand on Head) (f) Point Right (g) Point Left (h) Casual Stance (i) Attentive (j) Relax (k) Shrug. Pictures are used from *www.posemaniacs.com*

A. Recognized Human Postures

Postures are categorized mainly as sitting and standing. Overall 11 postures are recognized for standing and 8 postures have been recognized for sitting. Different postures recognized are crossed arms, open arms, think (hand on the head), think (hand on chin), pointing (with left hand), pointing (with right hand), standing/sitting normal, shrug, relax, casual posture and attentive posture. Figure 5 shows different postures for standing that are recognized by the system. Similar postures for sitting are also recognized by the system. Total of 10 subjects featured in the training stage. For each class and each subject, at least 300 instances are collected with a little bit of movement and varied styles.

TABLE I. Standing Postures and their Recognition Rates

	Postures	*Recog. Rate (%)*
1	Crossed Arms	100
2	Open Arms	100
3	Standing Normal	97.33
4	Think (Hand on the Head)	98.67
5	Think (Hand on the Chin)	95.1
6	Point with Left Hand	100
7	Point with Right Hand	100
8	Casual Stance	100
9	Standing Attentive	90.27
10	Shrug	100
11	Relax (Hands behind the neck)	100
	Average	**98.3**

B. Experimentation

There are generally two ways to conduct experiments. Experimentation or testing of the system can either be done on the testing dataset or testing can be done real-time directly on the ROBIN. We have conducted both these experiments in this work to evaluate the system. 25% of the dataset have been separated from the training data before training. This dataset serves as test dataset to evaluate the system. Since the recorded dataset has no false skeleton tracking, the system reports 99.4% recognition rate. This shows the potential of the approach when the provided dataset is accurate.

For the second experiment, ROBIN is used to recognize postures in real-time. Once ROBIN recognizes the posture, it indicates by saying the name of the posture. In order to avoid any bias, new subjects have been used to express postures in front of ROBIN. Subjects have been instructed in the start about the postures which ROBIN can recognize. However, the knowledge about performing each posture has not been shared with them in order to evaluate the system potential to generalize varied postures. Every subject performs each posture at least 30 times. Table I and Table II show the recognition rates of standing and sitting postures respectively.

TABLE II. Sitting Postures and their Recognition Rate

	Sitting Postures	*Recog. Rate (%)*
12	Sitting Normal	98.33
13	Crossed Arms	95.28
14	Think (Hand on the Head)	96
15	Think (Hand on the Chin)	94.5
16	Point with Left Hand	100
17	Point with Right Hand	100
18	Shrug	100
19	Relax (Hands behind the neck)	100
	Average	**98.01**

ISBN: 978-1-61208-616-3

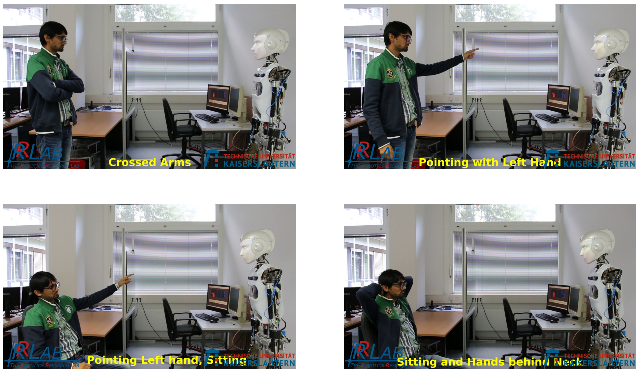

Figure 6. Subject is interacting with ROBIN using Postures.

C. Performance Evaluation

As shown in Table I and Table II, ROBIN is able to recognize human postures with an average accuracy of 98%. For standing postures, the recognition rate for each class is above 90%. Attentive posture reports low accuracy as compared to others because of the fact that the hands are too close to the body and therefore, the algorithm considers hands as part of the body for skeletal joints detection. Thinking postures are sometimes confused between each other and show recognition rates above 95%.

For sitting postures, it has been found out that when the person is sitting, the skeleton of whole body is not visible. In order to address this issue, ROBIN uses torso pitch angle to tilt its body in the front. In this way, the whole skeleton of human is visible. Due to sitting posture, sometimes human skeleton tracker does not work accurately to localize limbs and positions. Therefore, some of the postures show relatively less recognition rate than standing postures. Nevertheless, ROBIN is able to recognize human postures accurately in real-time with an accuracy of more than 98%. Since the system uses only depth data, issues regarding lighting condition, image resolution, texture variations are avoided. This enhances the accuracy considerably as compared to approaches using color image to recognize human postures. Figure 6 shows experimental environment where subject is interacting with ROBIN using postures.

V. Conclusion and Future Work

Identification of human postures is a complicated task based on the situation and interacting environment. Recognition of human postures has many applications in modern human-robot interaction developments. This can be applied for the purposes, e.g., natural interaction, gaming, developing assisted systems, surveillance systems, entertainment purposes and educational purposes. This paper presents an approach which uses RGB-D sensor for posture recognition. Depth information is used to extract joint positions. These joint positions are then converted into joint angles in order to make the system invariant to height, scale, position or physique of the person. Feature vectors are generated based on refined joint angles. SVM is used for classification of 19 different sitting and standing postures. System reports 100% recognition rate on a dataset with no false skeleton tracking and 98% when tested real-time in a cluttered and dynamic environment.

For future work, this approach can easily be extended for recognition of more postures. Additionally, using color image along with the depth can provide texture information, which can be utilized when the skeleton tracker does not work accurately.

References

[1] P. Noller, Nonverbal Communication in Close Relationships, 0th ed. SAGE Publications, Inc., 2006, pp. 403–421.

[2] B. D. Bruyn, “Review: The history of psychology: Fundamental questions,” Perception, vol. 32, no. 11, 2003, pp. 1409–1410.

[3] N. Dael, M. Mortillaro, and K. Scherer, “Emotion expression in body action and posture,” vol. 12, 11 2011, pp. 1085–101.

[4] J. Montepare, E. Koff, D. Zaitchik, and M. Albert, “The use of body movements and gestures as cues to emotions in younger and older adults,” Journal of Nonverbal Behavior, vol. 23, no. 2, Jun 1999, pp. 133–152.

[5] I. Rossberg-Gempton and G. D. Poole, “The effect of open and closed postures on pleasant and unpleasant emotions,” The Arts in Psychotherapy, vol. 20, no. 1, 1993, pp. 75 – 82, special Issue Research in the Creative Arts Therapies.

[6] S. Oosterwijk, M. Rotteveel, A. Fischer, and U. Hess, “Embodied emotion concepts: How generating words about pride and disappointment influences posture,” vol. 39, 04 2009, pp. 457 – 466.

[7] L. Al-Shawaf, D. Conroy-Beam, K. Asao, and D. M. Buss, “Human emotions: An evolutionary psychological perspective,” Emotion Review, vol. 8, no. 2, 2016, pp. 173–186.

[8] A. Xtion, “Xtion pro live — 3d sensor — asus global,” 2018, online; accessed 2018-03-12. [Online]. Available: https://www.asus.com/3D-Sensor/Xtion_PRO_LIVE/

[9] R. R. Lab, “Robin: Robot-human interaction,” 2018, online; accessed 2018-01-31. [Online]. Available: https://agrosy.informatik.uni-kl.de/robots/robin00/?L=1

[10] J. Wang et al., “Wearable sensor based human posture recognition,” in 2016 IEEE International Conference on Big Data (Big Data), Dec 2016, pp. 3432–3438.

[11] M. W. Lee and I. Cohen, “A model-based approach for estimating human 3d poses in static images,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 6, June 2006, pp. 905–916.

[12] A. Agarwal and B. Triggs, “3d human pose from silhouettes by relevance vector regression,” in Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004., vol. 2, June 2004, pp. 882–888 Vol.2.

[13] J. Malik and G. Mori, “Estimating human body configurations using shape context matching,” in Proceedings of the 7th European Conference on Computer Vision-Part III, ser. ECCV ’02. Springer-Verlag, 2002, pp. 666–680.

[14] S. Nirjon et al., “Kintense: A robust, accurate, real-time and evolving system for detecting aggressive actions from streaming 3d skeleton data,” in 2014 IEEE International Conference on Pervasive Computing and Communications (PerCom), March 2014, pp. 2–10.

[15] Z. Zhang et al., “A novel method for user-defined human posture recognition using kinect,” in 2014 7th International Congress on Image and Signal Processing, Oct 2014, pp. 736–740.

[16] I. Lillo, J. C. Niebles, and A. Soto, “Sparse composition of body poses and atomic actions for human activity recognition in rgb-d videos,” Image and Vision Computing, vol. 59, 2017, pp. 63 – 75.

[17] Y. L. Boureau, F. Bach, Y. LeCun, and J. Ponce, “Learning mid-level features for recognition,” in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 2010, pp. 2559–2566.

[18] J. C. Niebles, C.-W. Chen, and L. Fei-Fei, Modeling Temporal Structure of Decomposable Motion Segments for Activity Classification. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 392–405.

[19] S. Tong and E. Chang, “Support vector machine active learning for image retrieval,” in Proceedings of the Ninth ACM International Conference on Multimedia, ser. MULTIMEDIA ’01. New York, NY, USA: ACM, 2001, pp. 107–118.