This is the author's version of an article that has been published in IEEE RO-MAN 2012, DOI 10.1109/ROMAN.2012.6343882.
Changes were made to this version by the publisher prior to publication.
The final version of record is available at https://ieeexplore.ieee.org/document/6343882/

Affect Detection from Body Language during Social HRI

Derek McColl and Goldie Nejat, *Member, IEEE*

***Abstract*— In order for robots to effectively engage a person in bi-directional social human-robot interaction (HRI), they need to be able to perceive and respond appropriately to a person's affective state. It has been shown that body language is essential in effectively communicating human affect. In this paper, we present an automated real-time body language recognition and classification system, utilizing the Microsoft® Kinect™ sensor, that determines a person's affect in terms of their accessibility (i.e., openness and rapport) towards a robot during natural one-on-one interactions. Social HRI experiments are presented with our human-like robot Brian 2.0 and a comparison study between our proposed system and one developed with the Kinect™ body pose estimation algorithm verifies the performance of our affect classification system in HRI scenarios.**

I. INTRODUCTION

Social human-robot interaction (HRI) is an important subset of HRI that involves robots that are designed to communicate with humans through natural social cues. In particular, to be effective in social HRI, a robot must be able to perceive and interpret both verbal and nonverbal communication (i.e., body language, facial expressions and paralanguage).

It has been found that nonverbal behaviors communicate intent more effectively than verbal statements, particularly in communicating changes in affect [1]. Research to date has mainly concentrated on developing automated systems for interpreting affect from paralanguage and facial expressions, e.g., [2,3]. Thus far, few researchers have focused on automatically identifying affect from static body poses and postures [4,5]. In [4], joint angles of people playing a video game, obtained with a motion capture system, were manually segmented into poses corresponding to winning and losing scenarios. These poses were automatically classified into concentrating, defeated, frustrated, and triumphant affective states. In [5], pose information from a pressure sensitive chair was combined with facial features, skin conductance, and a pressure sensitive mouse to automatically determine if a child was becoming frustrated.

Our research concentrates on developing an automated affect classification system from the display of static body poses during one-on-one social HRI. It has been found that during interactions between two people, body language can be the most important contribution of information for the understanding of affective states [6]. Additionally, Mehrabian has also shown that body positioning relates the attitude of a communicator towards an addressee [7]. Hence, it is important that a robot be able to perceive and interpret body language during social HRI to more effectively engage a person with its own appropriate display of behavior.

Although several robots have been developed to understand human gestures as input commands, e.g. [8,9], they have not yet been designed to take into account static body poses to perceive and interpret a person's affective state during social HRI. Our objective is to develop a non-contact body language recognition and classification system for our socially interacting robot Brian 2.0, shown in Fig. 1, that can determine a person's affective state based on his/her static body poses. Brian 2.0 is designed as a non-contact socially interactive robot capable of task assistance. To effectively offer assistance, the robot must be able to both interpret and communicate information using natural non-verbal communication means such as body language.

Uniquely in this paper, we present an automated affect recognition and classification system using 2D and 3D sensory information provided by the Microsoft® Kinect™ sensor. The sensor is integrated onto Brian 2.0, Fig. 1, for perception during one-on-one social human-robot interactions. In our work, human affect is measured by the degree of accessibility (i.e., openness and rapport) of a person towards the robot. Research has shown that there is an important relationship between body language and a person's degree of accessibility. Specifically, the static body poses that a person displays during a one-on-one interaction scenario provides insight into the psychological state of that person [10]. Herein, we apply the Davis Nonverbal States Scale (DNSS), [10], to determine a person's degree of accessibility towards Brian 2.0 as determined from his/her static body poses during one-on-one interactions.

The paper is organized as follows. Section II describes the proposed automated body language recognition and classification system. Section III presents experiments that verify the integration of the proposed system into a social robot partaking in HRI scenarios. A detailed performance comparison study between our proposed 3D body language identification and classification method and the Kinect™ body pose estimation technique is also presented in Section III. The conclusions of the paper are presented in Section IV.

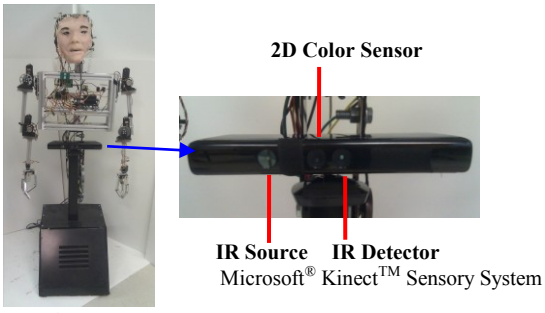


Figure 1. The Socially Assistive Robot Brian 2.0 and its Sensory System.

This research was partially supported by the Natural Sciences and Engineering Research Council of Canada and the Ontario Graduate Scholarship for Science and Technology.

D. McColl and G. Nejat are with the Autonomous Systems and Biomechatronics Laboratory in the Department of Mechanical and Industrial Engineering, University of Toronto, (e-mail: derek.mccoll@utoronto.ca and nejat@mie.utoronto.ca).

This is the author's version of an article that has been published in IEEE RO-MAN 2012, DOI 10.1109/ROMAN.2012.6343882.
Changes were made to this version by the publisher prior to publication.
The final version of record is available at https://ieeexplore.ieee.org/document/6343882/

II. Body Pose Estimation and Classification

In our work, body language is defined to represent static body poses that a person displays during social HRI in order to convey affect. Determining human body pose is a challenging task during social HRI because of the high dimensionality of the search space, large number of possible configurations and the sole use of on-board sensors on a mobile robot. In this paper, we present the development and implementation of a real-time sensor fusion technique utilizing the Kinect™ sensor to recognize and classify human body language during social HRI.

The work presented in this paper extends our previous research in static body pose recognition and classification [11,12]. In particular, this paper presents a robust fully automated person independent static body pose recognition and classification system utilizing the inexpensive Kinect™ sensor. We present the first use of the Kinect™ sensor for *affect classification* during one-on-one social HRI scenarios. 2D color and depth information is obtained from the Kinect™ sensor to first perform human body extraction to separate the depth and color information of the person from background information as well as segment individual body parts. These body parts are monitored to determine when a person is displaying a static pose. Once a static pose is identified, the segmented body parts are fit with ellipsoids. An ellipsoid human upper body model is generated using a reverse tree structure to represent static poses of a person. The ellipsoid model parameters are then used to classify a person's degree of accessibility towards the robot during interactions. Each step of this procedure is discussed in the following subsections.

A. Static Body Pose Definition

During social interaction, static body poses can provide information about a person's accessibility (rapport and openness) towards another person. The static body poses utilized in this work are derived from the Nonverbal Interaction States Analysis (NISA) of the Davis Nonverbal States Scale (DNSS) [10] and are defined as static poses that are held for at least four seconds. These static poses are utilized by NISA to determine how accessible a person is towards another person during interaction.

We utilize NISA's arrangement of trunk leans and orientations as well as arm positions to determine a person's accessibility level towards a robot. Namely, the upper trunk and lower trunk are each defined as: *Toward* (T) when each is oriented between 0^o to 3^o from the robot, *Neutral* (N) when each is oriented 3^o to 15^o from the robot, or *Away* (A) when each is oriented more than 15^o. Trunk lean is defined as *upright* when a person's shoulders are over the hips and *forward/back* when the shoulders are closer/farther than the hips in relation to the robot, right/left when the right/left shoulder is tilted past the right/left hip. The arm positions are defined at T when the arms are closer to the robot than the upper trunk, A when the arms are farther from the robot than the trunk or N when neither T or A.

B. Kinect™ Sensor

The Kinect™ sensor generates both depth and 2D color images at a resolution up to 640 x 480 pixels. The sensor was calibrated utilizing the Matlab® Calibration Toolbox [13] with a 3D checkerboard pattern. The sensor is mounted on Brian 2.0's mobile robotic platform, Fig. 1, and provides data for skin color identification and 3D segmentation of a person's upper body for 3D human body pose identification.

C. Human Body Extraction and Initialization

Depth-based foreground segmentation is performed to segment the person from the background depth data. Due to the one-to-one correspondence between the depth and 2D data, the background 2D data is also removed, e.g. Fig. 2(a). At the beginning of the interaction, an automated pose initialization step is performed, where the person is asked to stand facing the robot with arms hanging at his/her sides. Anthropometric information, [14], is utilized to estimate the locations of the waist and hips during initialization.

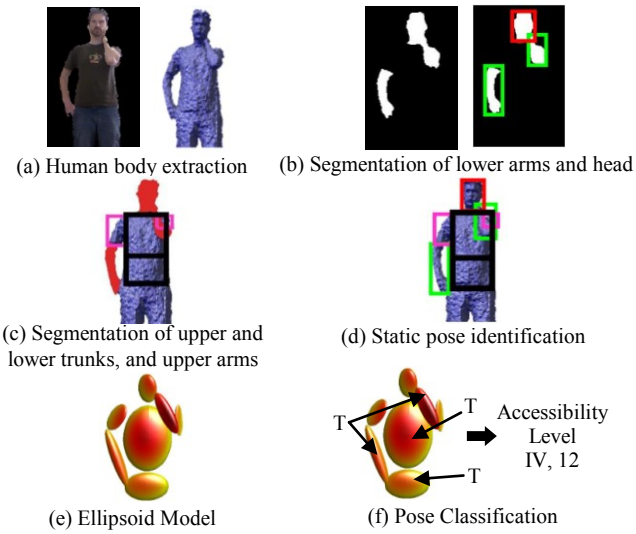


Figure 2. Body Pose Identification and Classification Procedure.

D. Human Body Part Segmentation

Body part segmentation is performed by first identifying the head and lower arms, then the upper and lower trunks, and lastly, the upper arms.

1) Lower Arms and Head Segmentation: The head and lower arms are identified utilizing skin color information from 2D color images. These body parts were chosen because they are readily exposed. The lower arms can be exposed by wearing short sleeves or rolling up long sleeves to the elbow. This is similar to numerous other HRI works that have clothing requirements for accurate body pose estimation or tracking, i.e., [8,9,15]. Skin regions are identified utilizing a pixel-by-pixel YCbCr color space range technique developed by Chai and Ngan [16].

Binary images of identified skin regions are utilized to determine regions that correspond to a head or lower arms, Fig. 2(b). In general, the number of skin regions that can represent these body parts can vary between 1 and 3, based on the number of occluded or touching body parts (since touching body parts appear as a single skin region).

Each skin region is identified utilizing three parameters, N_i, the number of pixels in the region, P_i, the number of pixels along the region's perimeter, and e_i, the eccentricity of the skin region. Utilizing these parameters, each skin region is classified into four cases, as shown in Table I: Case 1 – the head, Case 2 – a single lower arm, Case 3(a) – arms crossed,

This is the author's version of an article that has been published in IEEE RO-MAN 2012, DOI 10.1109/ROMAN.2012.6343882.
Changes were made to this version by the publisher prior to publication.
The final version of record is available at https://ieeexplore.ieee.org/document/6343882/

Case 3(b) – two skin regions touching (excluding crossed arms), and Case 4 – all three skin regions touching. Skin regions with $N_i < n$ pixels are classified as noise and removed from the binary image. Table I was developed based on the analysis of approximately 400 different NISA static body poses within the depth range of the Kinect™ sensor, i.e., 0.5-4m.

Skin regions identified as Cases 3 or 4 require separation into each body part to generate a full upper body model. Case 3(a) is separated into two lower arms by utilizing the major axis of the ellipse which was fit to the region to determine eccentricity as the separation line. Case 3(b) is separated into two body parts utilizing a technique that identifies a separation point between the body parts by analyzing paths of connected centroids of triangles generated by performing Delaunay triangulation on the region [11]. Fig. 2(b) shows the segmentation of an example pose for Case 3(b) with bounding boxes drawn around each segmented body part. It should be noted that although bounding boxes may overlap, no data is shared between body parts. Case 4 is separated by implementing the procedure for Case 3(b) twice to separate each of the three body parts. With the head and lower arms isolated, the remaining body parts can be segmented.

TABLE I: HEAD AND LOWER ARM CLASSIFICATION CASES

Case 1: Head	
$N_{H\min} < N_i < N_{H\max}$, where $N_{H\min}$ and $N_{H\max}$ are the minimum and maximum number of region pixels for a head.	(1)
$P_{H\min} < P_i < P_{H\max}$, where $P_{H\min}$ and $P_{H\max}$ are the minimum and maximum number of parameter pixels for a head.	(2)
$0 < e_i < e_{H\max}$, where $e_{H\max}$ is the maximum eccentricity for a head.	(3)
Case 2: Lower Arm	
$N_{A\min} < N_i < N_{H\min}$, where $N_{A\min}$ is the minimum number of pixels for a lower arm.	(4)
$P_{LA\min} < P_i < P_{LA\max}$, where $P_{LA\min} < P_{H\min}$ and $P_{LA\max} > P_{H\max}$. $P_{LA\min}$ and $P_{LA\max}$ are the minimum and maximum number of perimeter pixels for a lower arm.	(5)
$e_{H\max} < e_i < 1$.	(6)
Case 3(a): Crossed Arms	
$N_{H\max} < N_i < N_{CA\max}$, where $N_{CA\max}$ is the maximum number of region pixels for crossed arms.	(7)
$P_{LA\max} < P_i < P_{CA\max}$, where $P_{CA\max}$ is the maximum number of parameter pixels for crossed arms.	(8)
$e_{H\max} < e_i < e_{CA\max}$, where $e_{CA\max}$ is the maximum eccentricity for crossed arms.	(9)
Case 3(b): Two Arms Touching or Arm Touching Head	
$N_{CA\max} < N_i < N_{T\max}$, where $N_{T\max}$ is the maximum number of region pixels for two arms touching or one arm touching the head.	(10)
$P_{T\min} < P_i < P_{T\max}$, where $P_{T\min} >> P_{CA\max}$, and $P_{T\min}$ and $P_{T\max}$ are the minimum and maximum number of perimeter pixels for two arms touching or an arm touching the head.	(11)
$0 < e_i < 1$, this large range is required for e_i due to the wide variety of possible configurations of two arms touching or one arm touching the head.	(12)
Case 4: Both Lower Arms and Head Touching	
The skin region includes both lower arms and the head if it does not satisfy any of the previous cases.	

2) Segmentation of Upper and Lower Trunks, and Upper Arms: The upper trunk is defined as the region between the shoulders (minimum height of the region defined as the head) and the waist height (identified during initialization). The lower trunk is defined as the region between the waist and hip locations (identified during initialization). Removing the 3D data corresponding to the head, lower arms, and upper and lower trunks leaves the 3D data for only the upper arms and lower body (i.e., legs). The lower body region is easily identified as the largest remaining region that has the lowest centroid. This region is removed from the data, leaving behind the two remaining 3D data regions corresponding to the two upper arms. Example segmentations for the upper and lower trunks, and upper arms are presented in Fig. 2(c).

E. Static Pose Identification

The segmentation procedure described above is performed on every 10th frame of data captured by the Kinect™ sensor. Bounding boxes are formed around each body part, and the size and centroid of these bounding boxes are tracked to identify a static body pose. Image size normalization is applied to compare normalized centroids and bounding box sizes. If these parameters are within an allowable error of 2.5% when compared to the previous consecutive set of images, the current pose is defined to be the same as those in the previous frames. Example bounding boxes for body parts used for static pose identification are shown in Fig. 2(d). As previously mentioned, when a pose is held for 4 seconds it is defined as a static pose. Ellipsoids are then fit to the 3D data of each segmented body part of a static pose in order to generate a 3D human upper body model.

F. Reverse Tree Structure Ellipsoid Model

The 3D ellipsoid model of the static body pose is utilized with NISA to identify the accessibility of a person interacting with the robot. The 3D ellipsoids are fit to each of the seven segmented body parts. An example ellipsoid model is shown in Fig. 2(e). Ellipsoids are fit to the 3D data utilizing an iterative moment analysis technique similar to the technique presented in [17]. The ellipsoids for all the body parts are then connected together to form a full upper body ellipsoid model by applying a reverse tree structure. The head and lower arms are taken as the base of the reverse tree structure, to which the other body parts are connected. This reverse tree structure ensures that the appropriate orientations of the lower arms are maintained. The ellipsoid model can then be used to determine the degree of accessibility of a person towards the robot using the position accessibility scale of NISA, i.e. Fig. 2(f).

G. Accessibility Classification of Static Body Poses

The position accessibility scale consists of four levels, ranging from Level IV (most accessible) to Level I (least accessible) which have been defined through numerous clinical studies [10]. Each level is identified by the direction of trunk lean and the orientation patterns of the upper and lower trunks relative to the robot. These four levels are then subdivided into 3 sub-levels for finer scaling accessibility levels based on the T, N, or A arm patterns. Table II shows the combinations of upper and lower trunk orientations, trunk leans and arm patterns utilized to define each accessibility level.

This is the author's version of an article that has been published in IEEE RO-MAN 2012, DOI 10.1109/ROMAN.2012.6343882.
Changes were made to this version by the publisher prior to publication.
The final version of record is available at https://ieeexplore.ieee.org/document/6343882/

TABLE II: ACCESSIBILITY LEVELS

Trunk Orientation	Accessibility Level	Arm Orientation	Finer-Scaling
Upper/Lower trunk: T/N or N/T combined with upright or forward leans, T/T with all possible leans	IV	T	12
		N	11
		A	10
Upper/Lower trunk: T/N or N/T except positions that involve upright or forward leans	III	T	9
		N	8
		A	7
Upper/Lower trunk: N/N, A/N, N/A, T/A, A/T with all possible leans	II	T	6
		N	5
		A	4
Upper/Lower trunk: A/A with all possible leans	I	T	3
		N	2
		A	1

III. SOCIAL HRI EXPERIMENTS

Social HRI experiments were performed in our lab involving one-on-one interaction scenarios between a person and Brian 2.0. The objective of these experiments was to verify the performance of our Kinect™ sensor–based body language recognition and classification system. Eighteen students, ages 19 to 35, participated in the experiments, each naturally implementing a number of different static body poses during the interactions for detection, identification, and categorization into accessibility levels towards the robot. If participants were wearing long sleeves, they were asked to roll up their sleeves to their elbows. During the one-on-one social HRI interactions the participants stood an average of 1.4-1.6m from the robot, which is defined to be within the accepted social distance for human-human interactions [18]. Brian 2.0 was controlled in a Wizard of Oz fashion by a human operator in an isolated location from the robot and participant interaction area. The robot engaged each participant in the following four interaction stages:

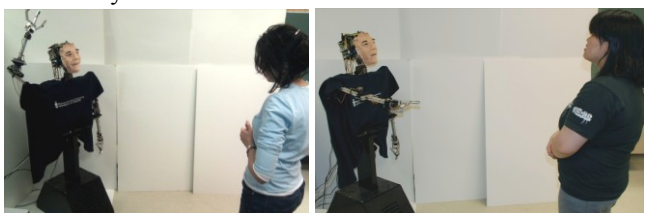
1) Introduction Stage: Brian 2.0 would introduce itself, explain its functionality and inquire about the person by asking questions. These questions included, for example, "Do you read? What is your favorite book?" Brian 2.0 would respond appropriately with general comments, such as "That sounds great." The robot would display body language such as waving and pointing and various facial expressions.

2) Instruction Stage: Brian 2.0 would provide the complete instructions on how to assemble a picnic table. These instructions were provided with a neutral facial expression.

3) Memory Stage: This stage consisted of a memory activity adapted from one of the questions from the Mini-Mental State Examination [19]. Brian 2.0 asked participants to remember a list of three objects, which at a future time it would ask the participant to recall. Prior to recall, the robot asked questions regarding the participant's childhood, focusing on his/her long term memories, such as "What is a happy memory from your childhood?" After approximately four minutes, the robot would ask the participant to state the three objects. The robot would offer congratulations if the participant answered correctly by smiling and saying "That is correct!" If the participant responded incorrectly, the robot would display a sad facial expression and say "Sorry that is wrong." The robot displayed various body gestures and facial expressions during this stage.

4) Repetitive Stage: Brian 2.0 would perform the same behavior repeatedly. In particular, it would ask a person to spell out the word "world", and then to spell it backwards. It would then repeat this behavior for 5 minutes. During this stage the robot kept its arms crossed with a neutral facial expression.

Figure 3 shows examples of robot behaviors during the interactions. The static body poses of the participants during these interactions were identified, and categorized into accessibility levels towards the robot.



(a) Brian 2.0 waving and smiling (b) Brian 2.0 pointing while talking

(c) Brian 2.0 with crossed arms while talking (d) Brian 2.0 with one hand on its head and a sad facial expression

Figure 3. Example behaviors of Brian 2.0 during interactions.

A. Results and Discussions

One hundred and two samples of different static poses displayed by the participants were recognized and classified with the proposed static body pose recognition and classification technique. Figure 4 presents eight typical static body poses displayed by the participants during the social HRI experiments.

The 2D and 3D sensory information, the proposed body part segmentation results and ellipsoid models are presented in Fig. 4 columns (i) to (iv), respectively. Segmentation results are shown as bounding boxes around each identified body part in column (iii). A red box corresponds to a head, green boxes - the lower arms, blue boxes - the upper and lower trunks, and magenta boxes - the upper arms. Although bounding boxes appear to overlap, no 3D data is shared between body parts. Figure 4 (a) and (g) have both the upper and lower trunks in a towards position. In (a) one hand is touching the chin while resting on the other arm and in (g) one hand is grasping the elbow of the other arm. For the pose in (b), the upper trunk is in a neutral position and the lower trunk is in a towards position with one hand touching the head and the other arm behind the trunk. For pose (c) the upper and lower trunks are in a neutral position and the arms are at the sides. For pose (d) upper and lower trunks are in towards position while the person is leaning forward with arms crossed. For pose (e) the upper trunk is in a neutral position and the lower trunk is in a towards position while leaning forward with both arms behind the trunks. For pose (f), both the upper and lower trunks are in an away position with arms at the sides. For pose (h), the upper trunk is in a neutral position and the lower trunk is in a towards position while leaning to the left with the arms crossed. Table III

This is the author's version of an article that has been published in IEEE RO-MAN 2012, DOI 10.1109/ROMAN.2012.6343882.
Changes were made to this version by the publisher prior to publication.
The final version of record is available at https://ieeexplore.ieee.org/document/6343882/

shows the accessibility levels determined from the ellipsoid models.

Figure 4. Experimental Results

As can be seen in Fig. 4, the ellipsoid models are excellent representations of the static body poses displayed by the participants. A human expert coder was used to determine the baseline static poses and accessibility levels for all the poses displayed in the experiments. Overall, 88% of the 102 ellipsoid models generated matched the baseline static poses. Currently, we estimate occluded body parts by utilizing ellipsoid parameters from previous frames and the present location of adjoining body parts. Occluded body parts are denoted by blue ellipsoids, as seen in Fig. 4(f) for the participant's right upper and lower arm. It was found that during interactions, the sleeves of the participants would move slightly higher or lower on their arms causing the estimated ellipsoids of the upper and lower arms to be longer or shorter. These body parts were always segmented separately, even with changes in sleeve position. It was found that 67% of the body poses during the interaction were classified as accessibility level IV, 1% as level III, 26% as level II and 6% as level I. These results signify that the participants, in general, were open to interacting with the robot. It should also be noted that all level I and level II poses were classified during the *Instruction* and *Repetitive* stages.

TABLE III: ACCESSIBILITY RESULTS

	Proposed Classification Method		Kinect™ body pose estimation	
Body Pose	***Accessibility Level***	***Finer-Scaling***	***Accessibility Level***	***Finer-Scaling***
Fig. 4(a)	IV	12	IV	12
Fig. 4(b)	IV	12	IV	11
Fig. 4(c)	II	5	II	5
Fig. 4(d)	IV	12	II	6
Fig. 4(e)	IV	10	II	4
Fig. 4(f)	I	2	I	2
Fig. 4(g)	IV	12	II	6
Fig. 4(h)	III	9	I	3

B. *Performance Comparison*

We conducted a two part performance comparison study between our proposed automated body language recognition and classification system and the Kinect™ body pose estimation algorithm in [20]. The first part of the comparison consisted of directly comparing the pose recognition results of the two techniques and the second part consisted of comparing the accessibility level results obtained from the two techniques with the expert coder's results.

1) Recognition Comparison: The Kinect™ body pose estimation technique identifies the 3D coordinates of 20 joints on a person's body from each frame of depth information [21]. Twelve of these points were utilized to generate the 3D skeleton model of the upper body needed for the accessibility scale, i.e., hands, elbows, shoulders, shoulder center, head, spine, hips and hip center. The same 102 poses that were identified above were used to generate the equivalent Kinect™ skeleton models. The skeleton models for the 8 poses in Fig. 4 are presented in column (v) as a direct comparison with our proposed technique. The accessibility levels generated utilizing the Kinect™ body pose estimation technique are also shown in Table III.

Forty-eight percent of the Kinect™ skeleton models did not accurately represent the arm poses of the participant as defined by the expert coder, especially when the arms are touching other body parts. This can be seen in Fig. 4(a) and (b) where the hand does not touch the head. In Fig. 4(d) and (h), the lower arms are not crossed appropriately and in Fig. 4(e) the arms are beside the body when they are actually behind the trunks. Lastly, in Fig. 4(g) the left arm is not holding the elbow of the right arm.

The random decision forest utilized by the Kinect™ body pose estimation technique was trained on a finite number of manually segmented sample depth images [20]. Hence, it is dependent on the training images used. Due to the very large number of possible poses and varying body shapes of individuals, the finite training set will not be able to cover all possibilities that exist. Furthermore, it has *not* been developed specifically for static pose recognition. This has resulted in the pose errors discussed above. On the other hand, the method proposed in this paper utilizes 2D color images, in addition to depth information, to identify the lower arms and head via skin color information. This allows the proposed method to accurately determine the pose of the

arms, even when touching other body parts. It is important to note that even though our proposed automated body language recognition technique requires an initialization pose, the Kinect™ body pose estimation algorithm requires a clear frontal view of the head and both shoulders, while the elbows need to be located lower than the shoulders and with no body parts touching the head in order to separate the depth information of a person from the background at the start of an interaction

2) Classification Comparison: The reliability of the accessibility level results from the proposed automated body language recognition and classification system, and the Kinect™ body pose estimation technique were compared with the accessibility results coded by the trained expert. The 2D color images of the 102 static body poses were provided to the coder to determine the participant's accessibility level. Table IV shows the results of the comparison between the coder and two automated accessibility classification techniques. The results of the comparison show that our proposed automated static body pose classification technique had 89% and 86% classification rates for the overall accessibility levels and finer-scaling, respectively. The Kinect™ body pose estimation technique resulted in 67% and 56% classification rates for the overall accessibility levels and finer-scaling. The difficulties with accurately identifying the arm poses, especially when touching other body parts, was the primary cause for the lower classification rates for the Kinect™ body pose estimation technique. Cohen's kappa was calculated to find that the strength of agreement between the proposed automated body pose recognition and classification technique and the Kinect™ body pose estimation technique with respect to the expert coder. Cohen's kappa was found to be 0.80 for our proposed approach, signifying a substantial strength of agreement and 0.45 for the Kinect™ body pose estimation technique, signifying a moderate strength of agreement [22].

TABLE IV: STATISTICS FOR PERFORMANCE COMPARISON

	Expert Coder	
Technique	***Accessibility Level Matches***	***Finer-Scaling Matches***
Our body language recognition and classification technique	89%	86%
Kinect™ body pose estimation	67%	56%

IV. CONCLUSION

In this paper, we present a unique real-time automated affect classification system for social HRI applications. The proposed body language recognition and classification technique utilizes the 2D color and depth information from the inexpensive Kinect™ sensor to perform full upper body part segmentation and 3D static body pose identification for automated classification of a person's accessibility level during one-on-one social HRI. Social HRI experiments as well as a detailed performance comparison study show the potential of integrating our body language recognition and accessibility level classification methodology into the human-like social robot Brian 2.0 in order for the robot to recognize and classify body language during one-on-one interactions. The results motivate future work to incorporate the body language recognition and classification system into the control architecture of Brian 2.0 to allow the robot to appropriately respond to a person's accessibility level.

REFERENCES

[1] S. Gong, P.W. McOwan and C. Shan, "Beyond Facial Expressions: Learning Human Emotion from Body Gestures," *British Mach. Vision Conf.*, pp. 1-10, 2007.

[2] M. El Ayadi, M.S. Kamel and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," *Pattern Recognition*, vol. 44, no. 3, pp. 572-587, 2011.

[3] Z. Zeng, M. Pantic, T.S. Huang, "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no.1, pp.39-58, 2009.

[4] A. Kleinsmith, N. Bianchi-Berthouze and A. Steed, "Automatic Recognition of Non-Acted Affective Postures," *IEEE Trans. Syst., Man, Cybern. Part B*, vol. 41, pp. 1027-1038, 2011.

[5] A. Kapoor, W. Burleson and R.W. Picard, "Automatic prediction of frustration" *Int. J. Human-Computer Studies*, vol. 65, no.8 , pp. 724-736, 2007.

[6] P. Ekman and V. Friesen, "Head and Body Cues in the Judgment of Emotion: A Reformulation," *Perceptual and Motor Skills*, vol. 24, no. 3, pp. 711-724, 1967.

[7] A. Mehrabian, "Significance of Posture and Position in the Communication of Attitude and Status Relationships," *Psychological Bulletin*, vol. 71, no. 5, pp. 359-372, 1969.

[8] B. Burger, I. Ferrane and F. Lerasle, "Multimodal Interaction Abilities for a Robot Companion," *Int. Conf. on Computer Vision Syst.*, pp. 549-558, 2008.

[9] V. Bonato et al., "A Real Time Gesture Recognition System for Mobile Robots," *Int. Conf. on Informatics in Control, Automation and Robotics*, pp. 207-214, 2004.

[10] M. Davis and D. Hadiks, "Non-verbal aspects of therapist attunement," *J. of Clinical Psychology*, vol. 50, no. 3, pp. 393-405, 1994.

[11] D. McColl, Z. Zhang, and G. Nejat, "Human Body Pose Interpretation and Classification for Social Human-Robot Interaction," *Int. J. Soc. Robot*, vol. 3, no.3, pp. 313-332, 2011.

[12] D. McColl and G. Nejat, "A Socially Assistive Robot That Can Interpret Human Body Language," *ASME Int. Design Eng. Tech. Conf.*, DETC2011-48031, 2011.

[13] Matlab Calibration Toolbox, Available HTTP: http://www.vision.caltech.edu/bouguetj/calib_doc/.

[14] M. Sanders and E. McCormick, *Human Factors in Engineering and Design. 7th Edition,* McGraw-Hill New York, 1993.

[15] G. Medioni et al. "Robust real-time vision for a personal service robot," *Comput. Vision and Image Understanding Archive*, vol. 108, no.1-2, pp. 196-203, 2007.

[16] D. Chai and K.N. Ngan, "Face Segmentation Using Skin-Color Map in Videophone Applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no .4, pp.551-564, 1999.

[17] G.K.M. Cheung, T. Kanade, J.Y. Bouguet, and M. Holler, "A Real Time System for Robust 3D Voxel Reconstruction of Human Motions," *IEEE Conf. on Comput. Vision and Pattern Recognition*, pp. 714-720, 2000.

[18] E.T. Hall. *The Hidden Dimension*, Doubleday, New York, NY, 1966.

[19] R. Crum, J. Anthony, S. Basset and M. Folstein, "Population-Based Norms for the Mini-Mental State Examination by Age and Education Level," *J. of the Amer. Medical Assoc.*, vol. 269, no.18, pp. 2386-2391, 1993.

[20] J. Shotton et al., "Real-Time Human Pose Recognition in Parts from Single Depth Images," *IEEE Conf. on Comput. Vision and Pattern Recognition (CVPR)*, pp. 1297-1304, 2011.

[21] Microsoft, "Kinect for Windows Programming Guide," Technical Document 2012, available from http://msdn.microsoft.com/en-us/library/hh855348.aspx.

[22] J.R. Landis and G.G. Koch, "The measurement of observer agreement for categorical data," *Biometrics,* vol. 33, no.1, pp. 159-174, 1977.