Human Attention Estimation by a Domestic Service Robot Using Upper Body Skeletal Information

H. P. Chapa Sirithunge, A. G. B. P. Jayasekara and D. P. Chandima
Robotics and Control Laboratory
Department of Electrical Engineering
University of Moratuwa
Moratuwa 10400, Sri Lanka
{chapa, buddhika and chandima}@elect.mrt.ac.lk

***Abstract*—Human-robot interaction is widely used for service applications in serving elderly and disabled population. A humanlike behavior is expected from these robots to achieve naturalness in interaction. Humanlike behavior is important before as well as after initiation of interaction with a human user. Basic cognitive skills are required for analyzing such a scenario involving humans. Many existing robots cannot interpret attention of its user before engaging with him. This paper presents a model to interpret the level of attention of the user when robot approaches him. The system extracts upper body information to approximate the level of attention of its user. This has separate functional units to extract and analyze information, estimate the attention level and finally to take decisions regarding interaction initiation with its user. It involves fuzzy evaluation of important parameters which affect the attention of a human. These parameters involve behavior of upper body of the person and using these, robot estimates the level of attention of its user without directly interacting. This model is implemented and tested in a domestic environment with users of a broad age gap. Implementation of this model and results of the experiment are presented.**

***Keywords*—Attention estimation; human robot interaction; nonverbal communication; interaction initiation**

I. Introduction

Service robots are available today, as a solution to the crisis of caretakers with the rise of disabled and elderly population [1]–[3]. These robots support daily tasks such as cleaning, cooking and taking care of health [4]–[6]. To achieve these tasks, service robots should possess cognitive and adaptation skills to interact with humans that they daily meet [7]–[10]. Robots equipped with the above qualities are accepted by its human user for interaction for a long duration.

Ability of interpreting the scenario associated with the user, at the same time the robot encounters a user is crucial in achieving an intelligent behavior. This helps the robot to determine when to interact and when not to. Furthermore, if an occasion is favorable for interaction, the robot determines what kind of interaction is required. This protects the naturalness and humanlikeness in interaction between the human and the robot. Such robots which resemble complex human behaviors are used as 'friends' or 'caretakers' in several systems such as [11]–[15]. However present robots often lack fluid interactive skills found during human-human interaction.

978-1-5386-1676-5/17/$31.00

Method used in [16], utilizes facial expressions and voice together with body postures to perceive the extent to which the user demands interaction from the robot. Voice is essentially an interaction demanding factor but it can only be analyzed after initiating a conversation. Systems [17], [18] evaluate postures and emotions during interaction to judge the likeliness of the user towards a friendly interaction. These systems help to adjust the behavior of robot during interaction, but it lacks the ability to perceive the user behavior before interaction.

A popular approach for interaction initiation can be seen in museum guide robots [19]. This system has the ability to select and approach visitors who demand attention from the robot. Spatial information, whole body orientation and mutual gaze times are considered as critical parameters to assess interaction demanding during this observation. However, there is bodily information that has more criticality during interaction, rather than orientation and spatial factors. Therefore this system may omit situations where body orientation and spatial factors are not favorable according to the system, but the person is still willing to interact. This situation is improved in [20] by increasing the number of parameters considered before interaction. One such important parameter is walking direction of the user. Method proposed in [21] estimates the degree of interest of the user towards an interaction with robot by measuring the attention and the distance to user and putting these inputs through a fuzzy logic based system. Here, attention of the user is related to the head tilt angle. Degree of user's interest is the output from this system. However, there are many facial factors which affect attention other than the head pose.

The above systems evaluate a limited number of parameters from the interaction scenario, to evaluate the possibility of initiating an interaction with a human user. Factors involving face, gestures and sometimes voice are involved when a human pays attention to another. Therefore, the presented system is developed such that it takes a number of important parameters that humans use to show attention towards someone nearby. These include major components such as gaze, changes in pose and friendly gestures to interpret attention. In addition, above major categories are sub divided into minor properties as well. For example, gaze is analyzed in 3 sub categories as gaze direction, gaze time and whether the gaze is returned to original position by the user. Therefore, this paper presents a model to interpret user's attention towards the existence of the robot in the surrounding, integrating major body and behavioral parameters. The system approaches analyzing data

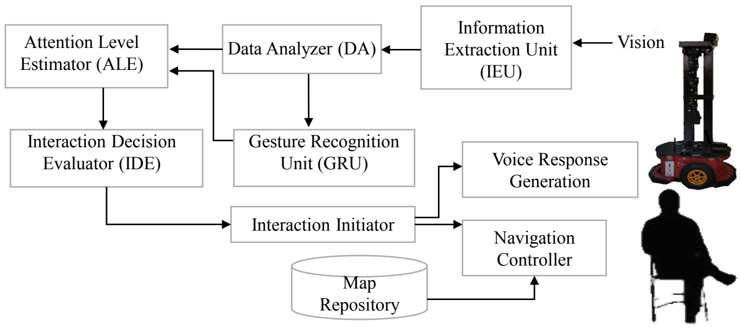

Fig. 1. Overview of the System

only after observation of a specific duration. Human user is tracked for his pose, movements and changes in upper body. Details and the implementation of the proposed system is presented in next sections.

II. System Overview

Overall system functionality is depicted in Fig. 1. The interaction between robot and its human user is interconnected through Voice Response Generation Module and the Vision System. Decisions regarding interaction with the identified human are evaluated from the modules that interconnect the above Voice and the Vision Systems. Information Extraction Unit (IEU) extracts upper body and facial information from the vision system and stores for a predefined t duration. After this time, Data Analyzer (DA) calculates parameters required as input for Attention Level Estimator (ALE) and Gesture Recognition Unit (GRU). GRU assess the changes in user's pose after noticing robot's presence. In addition, it identifies friendly gestures such as smile and waving hand. Calculated parameters from DA such as movement speeds and gesture times, and output from the GRU are fed into ALE to decide the level of attention of the person. Evaluation of input parameters is acquired from a fuzzy system in ALE. The type of interaction that has to be maintained with the particular user is decided by the Interaction Decision Evaluator (IDE), based on the output from ALE. Interaction Initiator is responsible for maintaining social distance between the robot and user, before interaction. Voice response is generated from Voice Response Generation Module after achieving social distance through Navigation Controller. Navigation maps of the environment are stored in Map Repository.

III. Identification of Gaze Parameters

The system recognizes gaze, friendly gestures and change in pose after the user has notified robot's presence. Out of these, gaze is analyzed under 3 sub categories as gaze time, gaze direction and gaze return. Gaze direction is zero if the user is looking directly at the camera mounted on robot. Gaze return is the extent to which the user turned his head back to original direction after looking at the robot. In the same way, friendly gestures are sub divided into happiness, gesture speed and gesture time wherein gesture speed and time are not involved with smile. How friendly gestures are evaluated is given in Section IV. Change in pose is evaluated separately without sub categories. These 3 are analyzed by a fuzzy system to evaluate the level of attention given by the user in each scenario.

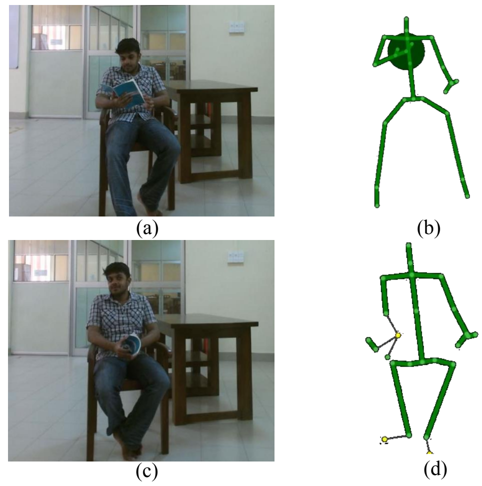

Fig. 2. Change of behavior encountered before and after approaching a human. In (a) Initially the user is engaged (b) skeletal which corresponds to that pose (c) change of pose when somebody reaches (d) Skeletal of the final pose

A. Determination of Gaze direction and Gaze time

Gaze is determined by the deviation of yaw angle of the user. This is taken from the point of view of camera mounted on robot. Yaw of the skeletal of a person tracked by the robot is denoted as θ_1.

$$|\theta_1| \leq \theta_{gaze\_lim} \tag{1}$$

where θ_{gaze_lim} is the angular limit for the gaze direction to be towards robot. If the condition in (1) is fulfilled, the system identifies the user to be looking at the robot.If the user satisfies the above requirement for t_{gaze} seconds, the gaze time will be t_{gaze}. The total duration system extracts information into IEU is t where $t_{gaze} \leq t$.

B. Return of Gaze

When the user is not interested in interacting with the robot, he will look away though he first looked at the robot.

$$|\theta_{1\_final} - \theta_{gaze\_lim}| \leq \Delta\theta_{lim} \tag{2}$$

Here, θ_{1_final} is the yaw at t^{th} second and $\Delta\theta_{lim}$ is the angular limit after which the gaze is decided to be returned. Above scenario is analyzed using the criteria given by (2). Gaze is not returned while the above criteria is fulfilled. In situations where the user looks at the robot and then looks

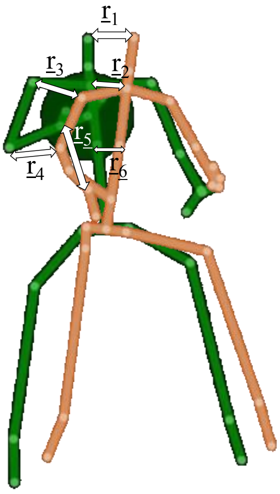

Fig. 3. Spatial vectors related to the body joints used to differentiate two poses of the same user. The initial pose and the final pose are shown in green and orange respectively.

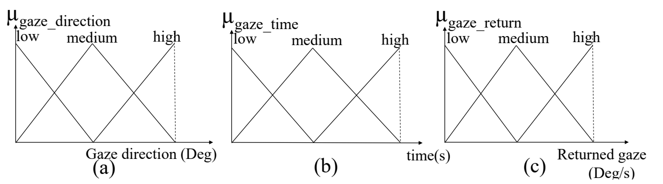

Fig. 4. Input membership functions of the gaze parameters (a) Gaze direction (b) Gaze time (c) Level of returned gaze Fuzzy labels: L, M, H, LD, MD, HD denote Low, Medium, High, Less deviated, Medium deviated and Highly deviated respectively.

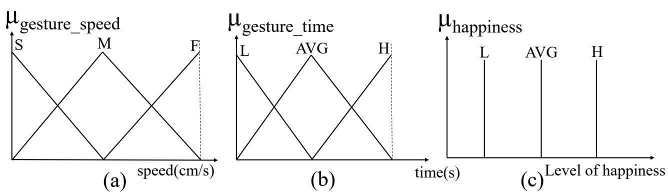

Fig. 5. Input membership functions of the gesture parameters (a) Gesture speed (b) Gesture time (c) Level of happiness Fuzzy labels: S, M and F for Slow, Medium and Fast respectively for Gesture Speed. L, AVG and H denotes Less, Average and High for Gesture time. L, AVG and H denote Low, Average and High respectively for Level of Happiness.

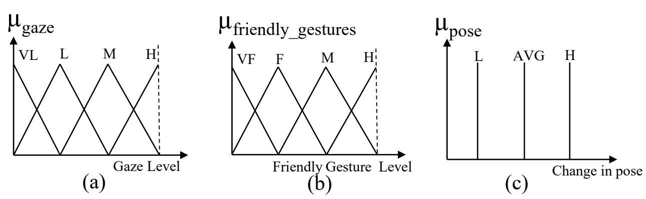

Fig. 6. (a) Output from Gaze parameters (b) Output from Friendly gestures (c) Change in Pose. These are used as the input parameters of second fuzzy evaluation Fuzzy labels: VL, L, M, H denote Very low, Low, Medium and Very high for Gaze Level. VF, F, M, H for Very few, Few, Medium and High in friendly gestures. L, AVG, H denote Less, Average and High for Change in Pose

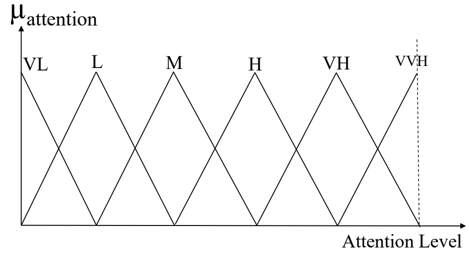

Fig. 7. Output Membership Functions of second fuzzy evaluation Fuzzy labels: VL, L, M, H, VH, VVH denote Very low, Low, Medium, Very high and Very Very High

away, many robots interpret such gaze as a favorable factor for interaction. But in real life, the condition is the opposite. Analyzing the above initial and final angular values, avoids a situation in which the robot tries to interact without the consent of its user. $|\theta_{1\_final} - \theta_{gaze\_lim}|$ decides the extent to which the user has turned his head back. All the 3 parameters, gaze direction, gaze time and return of gaze are analyzed by a fuzzy system to interpret gaze level. Gaze level is the extent to which the person's gaze is favorable from the point of view of an observer.

IV. Identification of Friendly Gestures

Smile and waving hand are used as the parameters which defines friendly gestures. Built in open/closed palm property in Kinect is used as one condition to look for waving hand. Built in 'smile' emotion recognition property in Kinect is used to detect smile.

$$y_{wrist} - y_{handtip} \geq y_{lim} \tag{3}$$

where y_{wrist} and $y_{handtip}$ are y coordinates of wrist joint and hand tip respectively as seen by the camera. y_{lim} is the angular limit after which the gesture will not be considered as waving hand. (3) must be fulfilled together with open/closed palm property for 'waving hand' gesture. Smile is detected using positions of the facial features in color and infra-red space, especially mouth left corner and mouth right corner.

$Gesture\_speed$ is evaluated by movement of open palm across the X-plane within a second of time. The speeds through out t is considered and average value is used as the input for fuzzy system. For example, the $gesture\_speed$ at the 1^{st} second out of t duration is shown in (4) below.

$$Gesture\_speed = \{x_{wrist}\}_{t=1} - \{x_{wrist}\}_{t=0} cm/s \tag{4}$$

$Gesture\_time$ is the duration at which hand waving is detected. $Happinenss$ is measured using smile and waving hand. $Happiness$ is considered 'High', if both smile and waving hand are present. $Happiness$ is 'Average' only one out of smile and waving hand is present. $Happiness$ is 'Low' if none of the two are present. All the 3 parameters, $gesture\_speed$, $gesture\_time$ and $happiness$ are required to evaluate gesture level of the user. Here, gesture level is the extent to which the person's friendly gestures are favorable for an interaction with an outsider.

TABLE I. RULE BASE TO EVALUATE GAZE LEVEL

	Gaze Direction								
	LD			MD			HD		
Gaze return	L	M	H	L	M	H	L	M	H
Gaze time									
L	-	L	VL	M	L	L	M	M	-
AVG	-	L	L	M	L	L	H	M	-
H	-	M	H	M	M	L	H	H	-

V. IDENTIFICATION OF CHANGES IN POSE

When a human notice another human expecting interaction with him, involuntarily he changes pose. This is analyzed as an input for the fuzzy system in ALE. This fuzzy system takes, changes in pose into account with gaze level and gesture level talked in the previous section. This scenario is illustrated in Fig. 2. Skeletal corresponding to the particular pose is shown on right of the image. Vectors shown in Fig. 3 are used to determine how the initial skeletal is changed from the final. IEU keeps track of the following vectors for further analysis in GRU. Vectors shown in the figure are expanded in (5)-(10) below.

$$\underline{r}_1 = \Delta x_{head}\underline{i} + \Delta y_{head}\underline{j} + \Delta z_{head}\underline{k} \tag{5}$$

$$\underline{r}_2 = \Delta x_{neck}\underline{i} + \Delta y_{neck}\underline{j} + \Delta z_{neck}\underline{k} \tag{6}$$

$$\underline{r}_3 = \Delta x_{shoulder}\underline{i} + \Delta y_{shoulder}\underline{j} + \Delta z_{shoulder}\underline{k} \tag{7}$$

$$\underline{r}_4 = \Delta x_{elbow}\underline{i} + \Delta y_{elbow}\underline{j} + \Delta z_{elbow}\underline{k} \tag{8}$$

$$\underline{r}_5 = \Delta x_{wrist}\underline{i} + \Delta y_{wrist}\underline{j} + \Delta z_{wrist}\underline{k} \tag{9}$$

$$\underline{r}_6 = \Delta x_{spine}\underline{i} + \Delta y_{spine}\underline{j} + \Delta z_{spine}\underline{k} \tag{10}$$

For further clarification, consider vector $\underline{r}_1$. If i, j, k components of the vector satisfy below conditions, a considerable change has been observed in user's pose. Here,Δx,Δy and Δz are differences between final and initial x, y and z coordinates of the particular vector.

$|\Delta x_{head}| \geq \Delta x_{head\_lim}$, $|\Delta y_{head}| \geq \Delta y_{head\_lim}$, $|\Delta z_{head}| \geq \Delta z_{head\_lim}$

The other 5 vectors are also analysed in the same way as above. Vectors are related to joints in the upper body which are most probably involved in pose changes. Number of vectors which satisfy the above condition is selected as the input for the fuzzy system. Angular and magnitude limits for condition checking in this research are derived through experimentation.

VI. ATTENTION EVALUATION

Impact of gaze, friendly gestures and change in pose towards the attention of the user is evaluated from a fuzzy system. Before evaluation of these 3 parameters, gaze and friendly gesture levels are evaluated by another fuzzy system. Inputs to evaluate $gaze\ level$ are illustrated in Fig. 4. Triangular membership functions are used for the ease of handling information as the system is involved with a number of parameters. In the same way, Fig. 5 shows the input parameters and membership functions used to determine the $friendly\ gesture\ level$. $Gaze\ level$ and $friendly\ gesture\ level$ are used with $change\ in\ pose$ as inputs for the second fuzzy system. Evaluation of this system is shown in Fig. 6.This system evaluates the overall behavior related to attention. Output membership functions of the system are illustrated in Fig. 7.

A. Interpretation of Fuzzy Parameters

$Gaze\ level$ shown in Fig. 6(a) is evaluated from the input parameters in Fig. 4. The rule base for this is shown in Table I. Some of the rules are omitted in actual implementation. For example, When the $Gaze\_direction$ is 'Low', $Gaze\_return$ automatically becomes 'Medium' or 'High'. Therefore it does not become 'Low' at the same time. When the $Gaze\_direction$ is 'High', $Gaze\_return$ does not become 'High' at the same time. Rule base for $Friendly\ Gesture\ Level$ evaluation is also performed in the same way. Rule base to evaluate $Attention\ Level$ is given in Table II. Boundaries for membership functions were determined through implementation.

B. Decision Making

Final decisions are taken by considering $Gaze\ level$, $Friendly\ gesture\ level$ and $Change\ in\ pose$ to evaluate $Attention\ level$. Input parameters of this fuzzy system are shown in Fig. 6 and the output membership functions are illustrated in Fig. 7. The type of interaction with the user may vary according to the $Attention\ level$ determined by the robot. Factors such as whether to reach the user or not, duration of conversation etc. can be determined from the measured $attention\ level$.

VII. RESULTS AND DISCUSSION

A. Research Platform

The proposed concept is implemented on MIRob platform attached with a Microsoft Kinect sensor [22]. Navigation maps required in motion are created with Mapper3 Basic software. Skeletal representation of the human body is acquired as 3D coordinates of feature points in Kinect sensor. The experiment was carried out in an artificially created domestic environment. Angular data is extracted by the system once every second. Color, depth, infrared and body streams of Kinect camera are used to recognize facial expressions and gaze parameters. (x,y) co-ordinates of facial points are projected to 2D color space to identify basic facial expressions. Smile is used as a friendly gesture and yaw angle of the head is used for the gaze direction.

B. Experiment and Results

The experiment is carried out with 11 persons in a broad age gap from 25 to 58 years (Mean of 35.4 and SD of 11.4). The robot was placed at several locations of a domestic environment and it was allowed to wander within the specified map. Once a body is tracked, the robot stops its motion and records extracted information for consecutive 8 seconds which is denoted by t in previous sections. Two occasions during the experiment are explained.

The initial human pose was as in Fig. 8(a) when the robot starts observing the person. The robot waits and observes the tracked skeletal to store information required for further analysis. System calculates parameters after 8 seconds of

TABLE II. Rule base for Attention level evaluation

	Gaze Level															
	VL				L				M				H			
Friendly gesture level	VF	F	M	H	VF	F	M	H	VF	F	M	H	VF	F	M	H
Change in pose																
L	VL	VL	L	M	L	L	M	H	L	L	M	H	M	M	H	VH
AVG	VL	VL	L	M	M	M	M	H	M	M	M	VH	M	M	H	VVH
H	VL	L	M	H	M	M	H	H	M	M	M	VH	M	H	VH	VVH

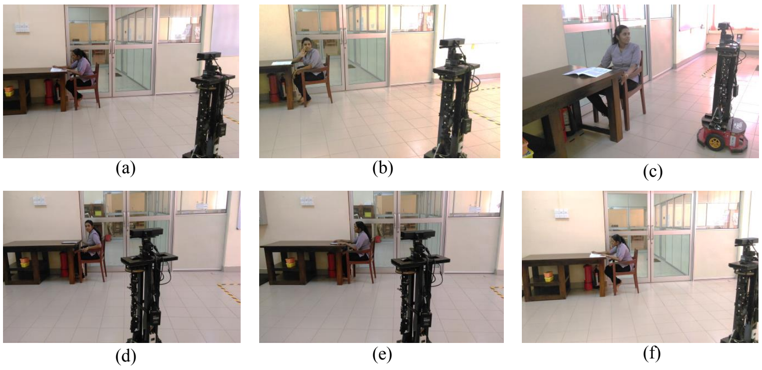

Fig. 8. Occasion 01 (a) User at the time of detection (b) User smiles and waves hand when she notices robot's presence (c) Robot approaches user Occasion 02 (d) User sees the robot (e) User looks away (f) Robot does not approach the user

observation. During implementation, $gaze\_direction$ is less deviated if gaze is smaller than 10°. $Gaze\_direction$ is highly deviated if gaze is greater than 60°. In this case, $gaze\_direction$ s observed by the robot was 25°. $Gaze\_time$ was recorded as 6s as the user started looking at the robot after a short while. Maximum $gaze\_time$ for the system is 8 s which has the fuzzy label 'High' while exact 'Medium' $gaze\_time$ is at 4 s. User waved hand until the robot started moving towards her. Therefore the gaze was not returned.

The moment user noticed robot's presence, her pose changed from Fig. 8(a) to Fig. 8(b). And she started waving hand while smiling. Therefore two conditions for $friendly\_gestures$ were also satisfied during the period of observation. Therefore $happiness$ of this scenario was 'High'. Evaluation of parameters according to the given fuzzy rules, defines a user of 'Very high' attention level which is a favorable situation for interaction. Therefore robot approaches user as in Fig. 8(c).

In the second situation, the user sees the robot as in Fig. 8(d), but she looks away after a short while as in Fig. 8(e). Therefore many parameters such as $gaze\_direction$, $gaze\_time$, $gesture\_speed$, $gesture\_time$ and $happiness$ are nullified. $Gesture\_return$ is 'High', nearly its maximum, 60°. Here, the gaze exceeds its limit of 60°. Therefore it is recorded as 60°. There was no $change\ in\ pose$ observed. Hence, after fuzzy evaluation as per the fuzzy rules given, attention level was 'Very Low'. Therefore the system decides this as an unfavorable situation for interaction. The robot did not approach the user. This can be seen in Fig. 8(f).

1) Results obtained from the proposed system : Results obtained by this method for several daily encountered tasks are given in Table III. Here, few tasks are considered due to the fact that the parameters considered in the system does not vary much depending on the task. Even so, evaluation of gestures, return of gaze can nullify the error occurred by not recognizing the tasks individually. Out of the 11 participants, the variation of their decisions for the same task if another human was engaged in each task is shown in the column named 'Humans with same decision'. In this scenario, the situation of the user depends on the task. Extracted information and calculated parameters are given. Approach decision to reach the user, making the mutual distance 1.5 m , is taken only when the interaction decision is 'yes'. Robot decides to interact only during 'Low', 'High' or 'Very high' attention levels are encountered. The system decision is compared with the decision of same group of participants. In this case, a participant is allowed to observe the human user behavior and his decision on initiating an interaction is recorded. In the same way, all the participants are allowed to take their own decisions regarding the suitability of an interaction during each task. During 'Sitting and making a call', human users did not try to interact, but the user's behavior made the system decide it as a favorable occasion for interaction initiation due to friendly gestures. This type of confusions are present due to the incapability of the system to identify tasks separately. During task 5, 'working on a table', due to the considerable change in pose, 4 users decided it as a favorable occasion for interaction if there's a necessity to interact. Otherwise the decision of the robot contradicts normal human behavior.

2) Comparison of results with a human study : A human study was conducted with the same 11 participants, to identify human tendency in the same scenario, if a human was in place of the robot. In other words, the human-robot interaction scenario was compared with a human-human scenario. This is shown under 'Humans with same decision' column and percentage accuracy under 'Accuracy' column in Table III. Despite the stated issues, the system performed satisfactorily during interaction scenarios.

VIII. Conclusion

A method has been introduced to interpret a human's attention towards the presence of a robot wherein the robot

TABLE III. EXPERIMENT RESULTS

Situation	Yaw (°)	Gaze return (°)	Gaze time (s)	Gestures	Speed (°/s)	Gesture time (s)	Changing vectors	Attention level	Approach Decision	Humans with same decision (out of 11 participants)	Accuracy
1 Sitting on a chair, task free	25	-	-	-	-	-	1	VL	-	11	100%
2 Sitting on a chair, task free	15	-	-	smile	-	-	2	L	1.5m	9	81.80%
3 Sitting, reading a book	0	45	2.1	-	-	-	0	L	-	10	90.90%
4 Working on a table	0	5	3.05	smile, waves hand	14	1.6	3	H	1.5m	10	90.90%
5 Working on a table	5	40	1.1	-	-	-	4	H	1.5m	4	36.40%
6 Sitting, in a call	0	40	2.3	smile, waves hand	26	1.8	2	M	1.5m	0	0.00%
7 Walking around	0	-	3.1	smile	-	-	6	H	1.5m	9	81.80%
8 Walking around	0	5	1.9	waves hand	24	2.1	6	VH		10	90.90%
9 Walking around	35	40	1.7	smile	-	-	6	L	-	11	100%
10 Walking away	45	-	-	-	-	-	6	VL	-	10	90.90%

approaches the user with the intention of initiating an interaction. The intention of the robot is starting a conversation with the user by analyzing posture and body movements related to joints. Before starting a conversation, the robot will be able to assess the attention given by the user towards robot. The major improvement of this system is making use of a nonverbal mechanism to identify favorable human behavior for interaction. Furthermore, the proposed system does not rely on verbal instructions from its user for interaction initiation. Experiments have been conducted to verify the performance of the system. Results show that the system is capable of understanding user situation without verbal communication with the user, as in many commonly used service robots. The system is able to perceive user attention towards robot through observation. In the future, it is expected to increase the number of parameters considered to evaluate human attention. Furthermore, attention comparison of multiple users is expected to be considered.

ACKNOWLEDGMENT

This work was supported by University of Moratuwa Senate Research Capital Grant SRC/CAP/17/03.

REFERENCES

[1] "World population ageing 2015," Population Division, Department of Economic and Social Affairs, United Nations, ST/ESA/SER.A/390, 2015.

[2] A. Koller and G.-J. M. Kruijff, "Talking robots with LEGO mindstorms," in *Proceedings of the 20th international conference on Computational Linguistics*. Association for Computational Linguistics, 2004, p. 336.

[3] J. M. Beer, C.-A. Smarr, T. L. Chen, A. Prakash, T. L. Mitzner, C. C. Kemp, and W. A. Rogers, "The domesticated robot: design guidelines for assisting older adults to age in place," in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. ACM, 2012, pp. 335–342.

[4] E. Prassler, A. Ritter, C. Schaeffer, and P. Fiorini, "A short history of cleaning robots," *Autonomous Robots*, vol. 9, no. 3, pp. 211–226, 2000.

[5] Y. Sugiura, D. Sakamoto, A. Withana, M. Inami, and T. Igarashi, "Cooking with robots: designing a household system working in open environments," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2010, pp. 2427–2430.

[6] R. M. Agrigoroaie and A. Tapus, "Developing a healthcare robot with personalized behaviors and social skills for the elderly," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*. IEEE Press, 2016, pp. 589–590.

[7] C. Plaisant, A. Druin, C. Lathan, K. Dakhane, K. Edwards, J. M. Vice, and J. Montemayor, "A storytelling robot for pediatric rehabilitation," in *Proceedings of the fourth international ACM conference on Assistive technologies*. ACM, 2000, pp. 50–55.

[8] M. M. de Graaf, S. B. Allouch, and J. A. van Dijk, "Long-Term Acceptance of Social Robots in Domestic Environments: In-sights from a Users Perspective," 2016.

[9] S. Restivo, "Bringing up and booting up: Social theory and the emergence of socially intelligent robots," in *Systems, Man, and Cybernetics, 2001 IEEE International Conference on*, vol. 4. IEEE, 2001, pp. 2110–2117.

[10] B. Robins, K. Dautenhahn, R. T. Boekhorst, and A. Billard, "Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills?" *Universal Access in the Information Society*, vol. 4, no. 2, pp. 105–120, Dec. 2005.

[11] S. Strohkorb and B. Scassellati, "Promoting Collaboration with Social Robots," in *The Eleventh ACM/IEEE International Conference on Human Robot Interation*. IEEE Press, 2016, pp. 639–640.

[12] G. Hoffman and C. Breazeal, "Robotic partners bodies and minds: An embodied approach to fluid human-robot collaboration," *Cognitive Robotics*, 2006.

[13] M. Heerink, B. Vanderborght, J. Broekens, and J. Alb-Canals, "New Friends: Social Robots in Therapy and Education," *International Journal of Social Robotics*, vol. 8, no. 4, pp. 443–444, Aug. 2016.

[14] B. Robins, K. Dautenhahn, R. T. Boekhorst, and A. Billard, "Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills?" *Universal Access in the Information Society*, vol. 4, no. 2, pp. 105–120, Dec. 2005.

[15] M. N. Nicolescu and M. J. Mataric, "Learning and interacting in human-robot domains," *IEEE Transactions on Systems, man, and Cybernetics-part A: Systems and Humans*, vol. 31, no. 5, pp. 419–430, 2001.

[16] H. Gunes and M. Piccardi, "Bi-modal emotion recognition from expressive face and body gestures," *Journal of Network and Computer Applications*, vol. 30, no. 4, pp. 1334–1345, Nov. 2007.

[17] M. P. Michalowski, S. Sabanovic, and R. Simmons, "A spatial model of engagement for a social robot," in *9th IEEE International Workshop on Advanced Motion Control, 2006*. IEEE, 2006, pp. 762–767.

[18] C. H. Chen and P. S.-P. Wang, Eds., *Handbook of pattern recognition and computer vision*, 3rd ed. River Edge, NJ: World Scientific, 2005.

[19] M. A. Yousuf, Y. Kobayashi, Y. Kuno, A. Yamazaki, and K. Yamazaki, "How to move towards visitors: A model for museum guide robots to initiate conversation," in *2013 IEEE RO-MAN*. IEEE, 2013, pp. 587–592.

[20] S. Satake, T. Kanda, D. F. Glas, M. Imai, H. Ishiguro, and N. Hagita, "How to approach humans?-strategies for social robots to initiate interaction," in *Human-Robot Interaction (HRI), 2009 4th ACM/IEEE International Conference on*. IEEE, 2009, pp. 109–116.

[21] E. Aguirre, M. Garcia-Silvente, A. Gonzlez, R. Pal, and R. Munyoz, "A fuzzy system for detection of interaction demanding and nodding assent based on stereo vision," *Journal of Physical Agents*, vol. 1, no. 1, pp. 15–26, 2007.

[22] M. A. V. J. Muthugala and A. G. B. P. Jayasekara, "MIRob: An intelligent service robot that learns from interactive discussions while handling uncertain information in user instructions," in *Moratuwa Engineering Research Conference (MERCon), 2016*. IEEE, 2016, pp. 397–402.