2012 7th International Conference on Electrical and Computer Engineering
20-22 December, 2012, Dhaka, Bangladesh

Robotic System for Making Eye Contact Pro-actively with Humans

Mohammed Moshiul Hoque,[1,*] Kausik Deb,[2]
[1]Graduate School of Science and Engineering, Saitama University
255 Shimo-Okubo, Sakura-Ku, Saitama 338-8570, Japan
[2]Dept. of Computer Science and Engineering, Chittagong University of Engineering and Technology
Chittagong-4349, Bangladesh
*moshiulh@yahoo.com

***Abstract*—Meeting eye contact is a most important prerequisite skill of a human to initiate an conversation with others. However, it is not easy task for a robot to meet eye contact with a human if they are not facing each other initially or the human is intensely engaged his/her task. If the robot would like to start communication with a particular person, it should turn its gaze to that person first. However, only such a turning action alone is not always be enough to set up eye contact. Sometimes, the robot should perform some strong actions so that it can capture the human's attention toward it. In this paper, we proposed a computational model for robots that can pro-actively captures human attention and makes eye contact with him/her. Evaluation experiment by using a robotic head reveals the effectiveness of the proposed model in different viewing infatuations.**

***Index Terms*—Human-robot interaction, attention attraction, eye contact, attentional focus.**

I. Introduction

Currently work in robotics is expanding from industrial robots to robots that are employed in the living environment. Human-robot interaction (HRI) is an interdisciplinary research field aimed at improving the interaction between human beings and robots and to develop robots that are capable of functioning effectively in real-world domains, working and collaborating with humans in their daily activities. For robots to be accepted into the real world, they must be capable to behaves in such a way that humans do with other humans. Although a number of significant challenges remained unsolved related to the social capabilities of robots, the robot that can pro-actively meets eye contact with human is also an important research issue in the realm of natural HRI.

Eye contact is a phenomenon that occurs when two people cross their gaze (i.e. looking at each other) which plays an important role in initiating an interaction and in regulating face-to-face communication [1]. Eye contact behavior is the basis of and developmental precursor to more complex gaze behaviors such as joint visual attention [2]. It is also a component of turn-taking that set the stage for language learning [3]. For any social interaction to be initiated and maintained, parties need to establish eye contact [4]. However, it is very difficult to establish such gaze behaviors for one person while the target person is not facing him/her or while target people are intensely attending his/her task.

A robot that naturally makes eye contact with human is one of its major capabilities to be implemented in social robots. Capturing attention and ensuring while capturing attention are the two important prerequisites for making an eye contact episode. Several previous HRI studies addressed the use of greeting behavior to initiate human-robot conversation [5], [6]. Some robots were equipped with the capability to encourage people to initiate interaction by offering cues such as approach direction [7], approach path [8], and standing position [9]. These studies assumed that the target person faces the robot and intends to talk to it; however, in actual practice this assumption may not always hold. Robots may wait for a person to initiate an interaction. Although such a passive attitude can work in some situations, many situations require a robot to employ a more active approach [10].

After capturing the attention of the intended recipient, the robot needs to make the person notice clearly that it is looking at none other than him/her after interpreting looking response. To solve this problem the robot should be able to display its awareness explicitly by some actions. Several robotic systems were incorporates gaze awareness functions by facial expression (i.e., smiling), head movement [11], and ear blinks [12]. To produce smiling expression, they used a flat screen monitor as the robot's head and display 3D computer graphics (CG) images. A flat screen is unnatural as a face. Moreover, these models used to produce the robot's gaze behavior are typically not reactive to the human partner's actions.

Situation where the human and the robot are not facing each other initially needs robots use a proactive approach to the intended human for making eye contact. This proactive nature is an important capabilities for robots that should be explored in the realm of HRI. This approach enables robots to help people who have potential needs and convey some information about an object or a particular direction that the human should focus. Moreover, to cope with the collaborative environment with human, the robot not only feedback against humans' needs but also convey its own intention toward the human. In summary, the major issues in our research are: (i) how can a robot use subtle cues to attract a human's attention (i.e., attention capture) if s/he is not facing to the robot, in other words, if the robot cannot capture his/her eyes or whole face due to the spatial arrangements of the person and the robot, and (ii) how robot ensure that the human is responding and how it tell when it has captured attention? To answer these issues we proposed an approach and we design a robotic head based on this that confirmed as effective to make eye contact with humans in experimental evaluation.

II. Our Approach

Humans usually turn their head first toward the person with whom they would like to communicate. If the target human does not respond, s/he tries with more strong signals

978-1-4673-1436-7/12/$31.00 ©2012 IEEE

(e.g., waving hand, shaking head, moving body etc.). Robots should use the same convention as humans in a natural HRI scenario. We determined to use head shaking if the robot cannot attract the target person's attention by basic head turning action because object motion is especially likely to draw attention [13]. Psychological evidence shows that the dynamic cues attract human attention irrespective to the level of cognitive load of a given task [14]. However, it may apparent that visual stimuli offered by the robot's nonverbal behaviors cannot affect a person if he/she is in a position where he/she cannot see the robot. In this situation, the use of touch or voice should be considered a last resort. Attention capture can produce observable behavioral responses such as eye, head movements, or body orientation, which often move together [15]. Therefore, if the person felt attracted by the robot, s/he will turn toward it. The robot should interpret the human looking response and display gaze-awareness which is an important behavior for humans to feel that the robot understands his/her attentional response. To ensure human response, the robot should be able to display its awareness explicitly by some actions. In this paper, we use eye blinking actions for the robot as its ensuring attention capture function.

Based on the above discussion, we can hypotheses that robots should perform two tasks consecutively: (i) attention capture, and (ii) ensuring attention capture for making eye contact pro-actively. Fig. 1 illustrates the conceptual process of attention attraction in terms of these tasks. To perform a successful eye contact episode, both a robot (R) and a human (H) need to show some explicit behaviors and to respond appropriately to them by communicative behaviors in each phase. That means, R and H performs a set of behaviors, $R = \{\phi, \psi\}$ and $H = \{\lambda, \delta\}$ respectively.

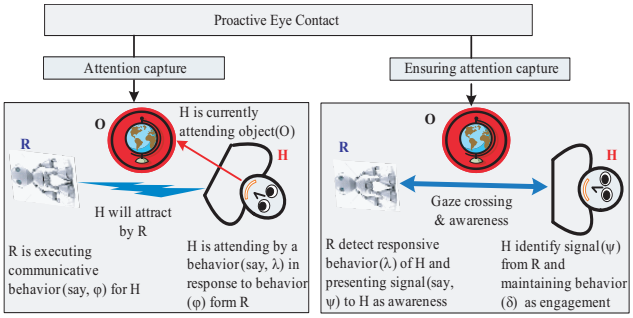

Fig. 1. Prerequisites of human attention capture process.

In this work, we apply a sets of behaviors of robot such as, $\phi = \{head\_turning, head\_shaking, reference\_terms\}$ in attention capture phase, and $\psi = \{eye\_blinks\}$ in ensuring attention capture phase. We are also expecting human behaviors such as, $\lambda = \{head \vee gaze \wedge body\_turn\_toward\_robot\}$ in attention capture phase, and $\delta = \{keep\_looking\_toward_robot\}$ in attention capture phase.

A. Behavioral Model of the Robot

The success of a particular action to attract human attention of a robot depends on the existing situation (i.e., direction of attention) as well as the nature of task that s/he is currently engaging. The robot can infer the current situation of the human by using the head information [16]. Fig. 2 illustrates how our robot system works in the situation where the human and the robot are not facing each other initially. An eye contact episode is initialize by detecting and tracking the human in interaction distance. Apply actions (head turning, head motions, and reference term) sequentially one after another. After each action, the robot waits (about 3 seconds), and checks whether or not the human is responding. If not, the robot selects the next available action until no actions are available (ending in failure to capture attention). If the human respond, robot detects his/her frontal face, and blinks its eyes to create awareness which ensure the attention capture process.

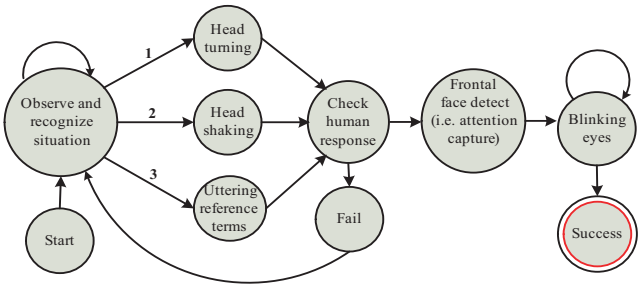

Fig. 2. Robot behaviors to capture the human attention. If first (1) option is fail, apply second (2) option. Third (3) option apply only when no option are available.

III. System Architecture

We have developed a robotic head for HRI experiments. Fig. 3 shows an overview of our robot head. The head consists of a spherical 3D mask, an LED projector (3M pocket projector, MPro150), laser range sensor (URG-04LX by Hokuyo Elec. Mach.), two USB cameras (Logicool Inc., Qcam), and a pan-tilt unit (Directed Perception Inc., PTUD46). The LED projector projects CG generated eyes on the mask. Thus, the head can show nonverbal behaviors by its head and eye movements including blinking. In the current system, one USB camera and laser sensor are putting on a tripod and placed at an appropriate position to observe human body and head. The proposed system consists of several software modules that are describes in the following.

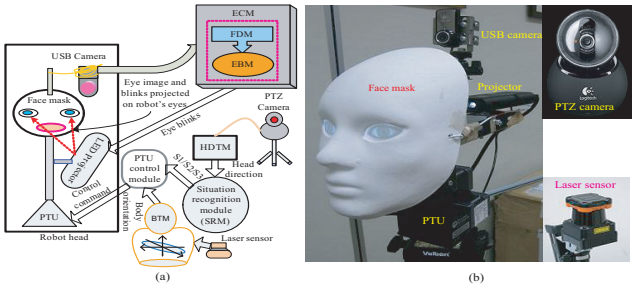

Fig. 3. A prototype of the robot head (rightmost photograph) that consists of four software modules (leftmost photograph).

Body Tracking Module (BTM): : After detecting current attentional focus of the target human, the robot should turn its head toward him/her for capturing attention. For turning head, the robot should know the location information of his/her gaze, or head or body. Our robotic agent continuously tracks the body direction of the target participant in real time using a laser range sensor and computes his/her body positions (x, y), directions (θ), and distance (D). Therefore, the robotic head turn its head toward the target person by using the information from the BTM.

A human body can be modeled as an ellipse (Fig. 4 (a)). We assume the coordinate system is represented with their X and Y axes aligned on the ground plane. Then, the human body model is consequently represented with center coordinates of ellipse $[x, y]$ and rotation of ellipse (θ). These parameters are estimated in each frame by the particle filter framework [17]. We assume that the laser range sensor is placed on the participant's shoulder level so that the contour of his/her shoulder can be observed. When the distance data which captured by the laser range sensor is mapped on the 2D image plane, the contour of the participant's shoulder is partially observed shown in Fig. 4 (b).

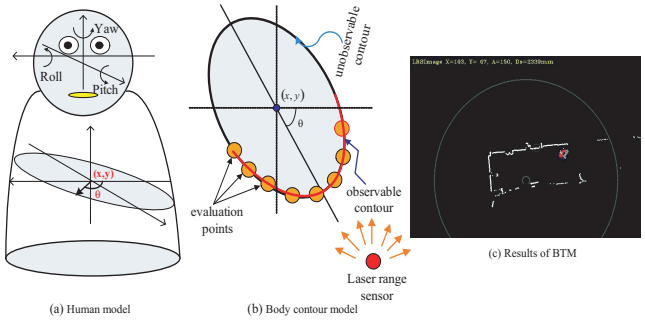

Fig. 4. Human body model.

The likelihood of each sample is evaluated the maximum distance between evaluation points and the nearest distance data using the Eq. 1.

$$\pi = exp\left(\frac{-d_{max}^2}{\sigma_d}\right) \tag{1}$$

where π is the likelihood score based on the laser image, d_{max} is the maximum distance between evaluation points and the nearest distance data. At each time instance, once the distance image is generated from the laser image, each distance d_n is easily obtained. σ_d is the variance derived from d_n. Evaluation procedures are repeated for each sample. We employ several points on the observable contour as the evaluation points to evaluate hypotheses in the particle filter framework. Selection of evaluation points can be performed by calculating the inner product of normal vectors on the contour and its position vector from laser range sensor. An example of the results of the BTM is shown in Fig. 4 (c).

Head Detection and Tracking Module (HDTM): To detect, track and computes direction of human face in real time (30 frame/sec), we use FaceAPI [18]. It gives the 3D-head coordinates (x, y, z) and corresponding rotational angles (α, β, γ) in radian. A snapshot of HDTM results has shown in Fig. 5 (a). The results of the HDTM send to the SRM to classify the current attentional direction (i.e. situation) of the target person.

Situation Recognition Module (SRM): : To recognize the situation (where the human is currently looking), we observe the head direction and body orientation estimated by HDTM and BTM respectively. By extrapolating information from the person's gaze (for central field of view, and peripheral field of view viewing situations) and body (for out of viewing situation), SRM determines the existing situation. The HDTM tracks within $\pm 90^0$ (right/left) only, therefore, while the human attend to OFOV situation, the system losses head information, in that case, the robot recognize the current situation based on the body information (laser sensor can tracks up to 270 degrees). From the results of tracking modules, the system recognizes the three viewing situations in terms of yaw (α), pitch (β) movements of head and/or body direction (θ) respectively using a set of predefined rules. For example, if the current head direction (of human with respect to the robot) within $-10^0 \leq \alpha \leq +10^0$ and $-10^0 \leq \beta \leq +10^0$ and remains 30 frames in the same direction, system recognized as the central field of view situation. In each rule, we set the values for yaw, pitch and body directions by observing several experimental trials.

Eye Contact Module (ECM): The ECM mainly consists of two sub modules: the face detection module (FDM), and the eye blinking module (EBM). The robot considers that the human has responded against the robot actions if s/he looks at the robot within expected times. In that case, FDM uses the image of the forehead camera to detect his/her frontal face (Fig. 5 (a)). We use the face detector based on AdaBoost classifier and Haar-like features [19]. After face detection, FDM sends the results to EBM for exhibiting eye blinks. Since the eyes are CG images, the robot can easily blink the eyes in response to the human's gazing at it. Figs. 5 (b)-(g)) illustrates some screenshots of eye behaviors of the robot.

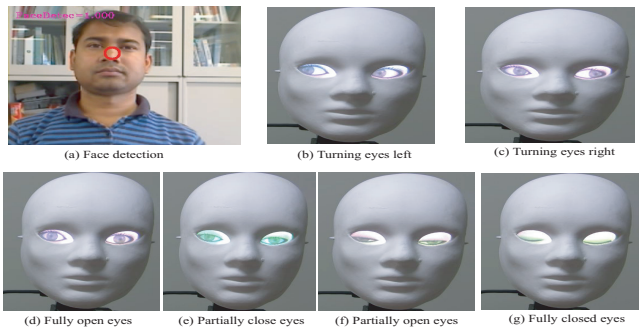

Fig. 5. Results of ECM.

IV. Evaluation Experiment

To verify the effectiveness of our proposed model, we conducted an experiment with a total of 48 graduate students of Saitama University, Japan. Their ages ranging from 22 years to 34 years. Participants were randomly assigned into three groups according to the viewing situations (central field of view, peripheral field of view, and out of filed of view).

A. Design and Procedures

We hanged five paintings (P1 to P5) on the wall at the same height and asked the participants to watch the paintings. If only a robot exists in the room, participants may be attracted by the robot even though it does not perform any action. Thus we prepared two robots to reduce the self attention rate. Both were the same in appearance. One was (i) Moving robot (MR). Initially MR is static and is looking in a direction not toward the human face. The second was (ii) Static robot (SR). It is stationary all times (i.e., does not perform any head movements) and is looking forward direction with blinking. The MR was placed at P11 (for central field of view (CFOV), and peripheral field of view (PFOV) situations) and P12 (for out of field of view (OFOV) situation) whereas the SR was placed at the right of the rightmost painting. Fig. 6 shows the experimental settings. We programmed the moving robot in two ways. (i) Proposed robot: It was behaved as described in

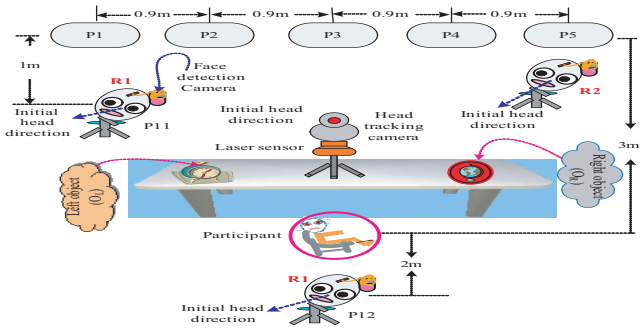

Fig. 6. Experimental set up.

section II. (ii) Conventional robot: Attention capture behaviors were remained the same as proposed robot but it did not displayed any blinking action after capturing people attention. Each group interacted with both robots one after another. All sessions were videotaped by two video cameras. Fig. 7 shows some experimental scenes.

(a) The participant is watching paintings (b) The participant has attracted and looking at MR

Fig. 7. Experimental scenes.

V. Evaluation Results

To evaluate the system, we measured two items: success rate and impression. To measure the success rate, we observed the number of participant's looking at the robot (N_L) against its total number of attempts (N_A), and number of times that it generates blinking (N_B). After interacting with robots, participants rated their feeling of making eye contact based on the 7-point Likert scale. Table I summarizes quantitative measures in three viewing situations for the proposed robot.

TABLE I
Success rate in diffract viewing situations

Situations	N_L			N_B	Success Rate
	HT	HS	RT	EB	%
CFOV ($N_A = 16$)	15	01	0	16	100
PFOV ($N_A = 16$)	02	12	02	14	87.5
OFOV ($N_A = 16$)	0	00	15	13	81.25

Results indicates that only head turning (HT) is enough when the robot presents in the participant's central field of view (CFOV), but the robot should be used more stronger action such as head shaking (HS) when it captured in in the participant's peripheral field of view (PFOV). Moreover, in out of field of view (OFOV) situation, any kind of motions did not effective, and in that case, only voice signal can be capture humans' attention effectively. After capturing the attention, robot will shows the eye blinking behaviors to ensure the attention capture process. Therefore, the success rate of the robot depends on the attention captures as well as ensuring attention capture. These results also shows that among 16 participants the robot can captures the all participants attention in CFOV and PFOV, and 15 participant's attention in OFOV situations. The robot proceeded to the next ensuring attention capture step for these successful cases. For all cases (16 out of 16, 100%), the proposed robot gave eye blinks and participants were looked at it during blinking in CFOV. However, the robot displayed blinks for 87% of times in PFOV, and 81% of times in OFOV respectively. In some cases, the robot did not start blinking due to failure of recognizing frontal face.

Subjective evaluation of participants' average impression shows that participants rated the proposed robot ($M = 5.56, SD = 0.78$) more that the conventional robot ($M = 3.14, SD = 0.55$). We Compared the 16 resultant pairs for each situations using t-test. The results shows a significant differences between two robots in CFOV ($t(15) = 11.2, p < 0.01$), PFOV ($t(15) = 6.3, p < 0.01$), and OFOV ($t(15) = 8.3, p < 0.01$) situations respectively.

VI. Conclusions

The primary focus of our work is to develop a model for social robots that would make eye contact with human when it intended. We have shown that our proposed approach is functioning to capture human attention pro-actively for making eye contact. We have shown that eye blinking in response to the human is effective to make a feel that the robot can understand his/her response. We have considered a particular scenario in this paper. If a person is intensely paying attention to his/her work, the robot needs to use some other actions such as waving, or touch. These are left for future work.

References

[1] M. Argyle, *Bodily Communication*, London: Routledge, 1988.
[2] T. Farroni and E. M. Mansfield and M. H. J. Carlo Lai, "Infant perceiving and acting on the eyes: Tests of an evolutionary hypothesis", *J. of Exp. Child Psy.*, vol. 85, pp. 199-212, July, 2003.
[3] C. Trevarthen and K. J. Aitken, "Infant intersubjectivity: Research, theory, and clinical applications", *J. of Child Psycho. & Psychiatry.*, vol. 42, pp. 3-48, Jan., 2001.
[4] E. Goffman, *Behavior in Public Places: Notes on the Social Organization of Gatherings*, 1st ed. New York: The Free Press, 1963.
[5] K. Hayashi et al., "Humanoid robots as a passive-social medium-A field experiment at a train station", *Proc. HRI'07*, 2007, pp. 137-144.
[6] M. P. Michalowski, S. Sabanovic, R. Simmons, "A spatial model of engagement for a social robot", *Proc. AMC'06*, 2006, pp. 762-767.
[7] K. Dautenhahn, M. Walters, S. Woods, K. L. Koay, C. L. Nehaniv, "How may I serve you?: A robot companion approaching a seated person in a helping context, *Porc. HRI'06*, 2006, pp. 172-179.
[8] C. Shi, M. Shimada, T. Kanda, H. Ishiguro, N. Hagita, "Spatial formation model for initiating conversation, *Proc. RSS'11*, 2011.
[9] F. Yamaoka et al., A Model of proximity control for information presenting robot, *IEEE Tran. on Robo.*, vol. 26, pp. 187-195, Feb., 2010.
[10] S. Satake et a., "How to approach humans? Strategies for social robots to initiate interaction, *Proc. HRI'09*, 2009, pp. 109-116.
[11] Y. Yoshikawa, K. Shinozawa, H. Ishiguro, N. Hagita, and T. Miyamoto, "Responsive robot gaze to interaction partner", *Proc. RSS'06*, 2006.
[12] C. M. Huang, A. L. Thomaz, "Effects of responding to, initiating and ensuring joint attention in human-robot interaction", *Proc. Ro-man'11*, 2011, pp. 65-71.
[13] W. James, *The Principle of Psychology*, New York: Dover, 1950.
[14] S. L. Franconeri, D. L. Simons, "Moving and looming stimuli capture attention", *J. Per. & Psychopsy.*, vol. 65, pp. 999-1010, Oct., 2003.
[15] C. Peters, "Direction of attention perception for conversation initiation in virtual environments", *Int. Vir. Age.*, vol. 3661, pp. 215-228, 2005.
[16] M. Argyle and M. Cook, *Gaze and Mutual Gaze*, Oxford: Cambridge University Press, 1976.
[17] M. Isard, A. Blake, "Condensation-conditional density propagation for visual tracking", *Int. J. of Com. Vis.*, vol. 29, pp. 5-28, 1998.
[18] (2010) FaceAPI. Available: http://www.faceapi.com
[19] G. Bradsky et al., "Learning based computer vision with Intel's open computer vision library, *J. Intel Tech.*, vol. 9, 2005, pp. 119-130.