A Robotic Framework for Shifting the Target Human's Attention in Multi-Party Setting

Mohammed Moshiul Hoque
Chittagong University of Engineering & Technology, Chittagong, Bangladesh
moshiul_240@cuet.ac.bd

Yoshinori Kobayashi, Yoshinori Kuno
Saitama University, Saitama, Japan
yosinori@cv.ics.saitama-u.ac.jp
kuno@cv.ics.saitama-u.ac.jp

Abstract— **It is a major challenge in HRI to design a robotic agent that is able to direct its partner's attention from his/her existing attentional focus towards an intended direction. For this purpose, the agent may first turn its gaze to him/her in order to set up eye contact. However, such a turning action of the agent may not in itself be sufficient to establish eye contact with its partner in all cases, especially when the agent and the its partner are not facing each other or the partner is intensely engaged in a task. This paper focuses on designing a robotic framework to shift the target human's attentional focus toward the robot's intended direction from multiple humans. For this purpose, we proposed a conceptual framework with three phases: capturing attention, making eye contact, and shifting attention. We conducted an experiment to validate our model in HRI scenarios in which two participant interacted in a session at a time. One of them interacted as a target and other as a non-target. Experimental results with twenty participants shows the effectiveness of the proposed framework.**

Keywords: human-robot interaction; capturing attention; eye contact; shifting attention

I. Introduction

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 environments, working and collaborating with humans in their daily activities. In the future, the robots would be coexist into the human world. 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 shifting a particular human's attention from multiple humans is also an important research issue in the realm of natural HRI. In this paper we focus on such a social function such as, *shifting the target human's attention*.

Attention shift capabilities, defined as reorienting or reallocating attention to a target because it is the object of another person's attention. Shifting attention of the intended human when s/he is involving her/his task is one of the fundamental skills in human social interaction and cognition. Moreover, it play a critical role in autonomous mental development such as regulating turn taking [1], exhibiting one's attention [2], and so on. Attention shift is the foundational mechanism by which independent agents coordinate their activity to develop more complex interactions, and it supports inferences about others agents' current and future activity, both overt and covert. Shifting of attention is necessary because it allows us to redirect attention to aspects of the environment we want to focus on, and subsequently process [3]. However, it is difficult for the robot to shift someone's attention from his/her current attentional focus to another, when s/he is not facing toward it or while s/he is intensely attending at his/her work. In that case, the robot should capture the target human's attention first and then establish eye contact with him/her. The robot's capability to shift people attention can be used is an invitation service or help to find a particular item. Such as a robot pro-actively invites people and shift their attention to visit a particular product. As an another example consider a scenario, a person is searching a particular item but he/she failed to locate this, in that case the robot understand his/her situation, pro-actively captures his/her attention for meeting eye contact, and then it help him/her to find this particular item by using its attention shift behaviors.

The success of a robot to control a target human's attention depends on the number of factors including target participant's existing situation, level of attention needed his/her current task, distance between the robot and the target human, actions played by the robot, and so on. However, it is difficult to consider all factors and actions together in early stages of research. Thus, as a first step, we limit our scope of research as follows. How can the robot shift a target human's attention alone among multiple humans from his/her current attentional direction toward its goal direction when s/he is not looking toward the robot initially or s/he is engaging such a task that does not absorb much attention to perform. We show that capturing attention, making eye contact, and shifting attention behaviors of the robot can shift the target human's attention toward the robot's goal direction in a experimental evaluation.

II. Related Work

In robotics, attention shift studies have recently been receiving increased attention [4]. Staudte and Crocker [5] also showed that people follow and use robot gaze, similar to human gaze, and they find clear evidence for gaze mediated joint attention. In HRI, the attention shift capabilities were designed from the human's perspective. That means, the human initiates the attention shift request by shifting his/her head, gaze, or pointing toward an object. The robot estimate his/her looking direction and moves its head to the same

978-1-4799-5180-2/14/$31.00 ©2014 IEEE

direction. However, to generate the natural HRI, the detection of gaze direction may be considered from both perspectives: the robot reading gaze direction of the human partner and the human partner reading robot's gaze. We review the recent progresses on the attention shift capability of robots on these two aspects in the following sections.

1) Human Initiative Approach: Most works have considered that the human initiates the attention shift request first using social cues and robot directs its head toward the directed cues. Nagai et al. [6] proposed model reflected the process that an infant acquires attention shift through interacting with a caregiver. Carlson and Triesch [7] suggested a computation model of emergence of gaze following based on a reinforcement learning method. Their result showed that gaze following can be learned in the context of proper interaction between the proposed mechanism. For humans, following pointing gestures of other humans is an important ability to jointly focus their attention on objects of interest. Approaches to endow robots with a similar capability were proposed by Heidemann et al. [8], Schauerte et al. [9], and Nagi [10]. They analyze the direction of pointing finger and fuse this top-down information with the bottom-up saliency of objects.

Sumioka et al. [11] incorporates motion cues of participant's to shift the robot's gaze toward the object. Some robots can follow people's attentional targets by means of monitoring their gaze-direction [12] and face direction [13]. In [14], Marin-Urias et al. implemented joint attention using a geometric reasoning mechanism based on mental rotation and perspective taking concepts. Attention shift behavior involves nonverbal social cues, such as eye gaze, which acts as transparent communication. Some robotic agents shift its attention toward the human's attended object in several ways including gaze turn [15], head and face orientation [16], pointing gestures [17], [18], and body pose [19].

The ability to detect the direction of an agent's attentional focus need to be present for both interacting partner, and hence it is not only important that the robot can interpret human attentional focus, but also that a human can perceive where a robotic partner is looking. However, all previous works focused on reading the attentional focus of human by robots. Moreover, Most of these assumed that a human faces to the robot when their interaction begins but this assumption may not be practical in natural HRI cases.

2) Robot Initiative Approach: For natural HRI, the robot should be able to accurately display its attentional focus toward a particular direction or object that can be correctly interpreted by human users. Some works addressed the aspects of initiating attention shift from the robot. Hunag et al. [20] focused on initiating the attention shift of the human partner via pointing and voice behaviors of the robot. Mutlu et al. [21] uses the gaze direction of the robot to indicate an object and the human partner is trying to interpret its gaze directed object. These works also assumed that the robot and the human are facing each other while interaction begins which may not be practical in natural HRI scenarios. Our previous work [22], [23] also described the attention shift mechanisms for robot initiated approach. However, we considered only the situations with a single participant interacted with the robot. Thus, it is remained unexplore how the robot shifts the particular person's attention toward its goal direction from multi-person scenarios. In this paper, we present an approach to attention shift that can selectively capture a particular person's attention in multi-party setting, establish eye contact with him/her, and shift his/her attention toward the robot's intended direction.

III. Hypotheses in Shifting Attention

Attention is the cognitive process of selectively concentrating on one aspect of the environment while ignoring other things [24]. Without information from the eyes, cues such as head orientation, body posture, and pointing gestures might also indicate direction of attention [25]. Although eyes are the primary source of information about the direction of attention, there are plenty of situations where determining the looker's eyes direction is impossible or infeasible. Specially, when people are at greater distances or people are not facing each other, head orientation becomes a stronger cue than information from the eyes in determining direction of attention. Thus, in attention shifting process, we would like to use the information of head direction to monitor the participant's visual focus of attention. A canonical attention shift episode, between the robot and the target participant, can be described in three major steps: capturing attention, making eye contact, and shifting attention. Fig. 1 shows a simple depiction of the attention shift process.

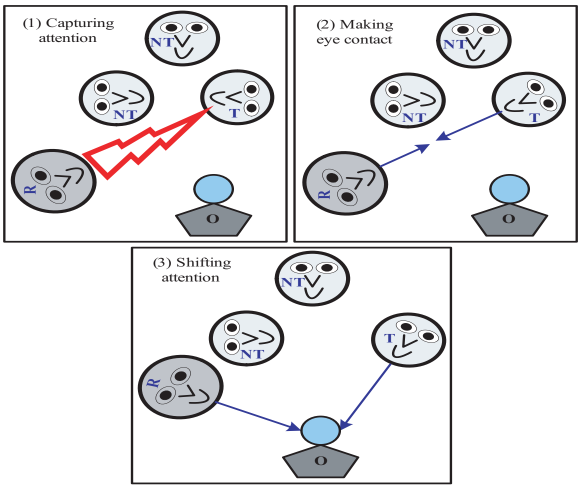

Fig. 1. Attention shift process in terms of three steps. Direction of arrows indicates the agent's attentional focus. Symbols (R, T, NT, and O) represent the robot, the target human, the non-target human and the object, respectively.

- **Capturing attention**: This refers to one's ability to capture another's attention for establishing a communication channel trough eye contact. The robot should behaves in such a way that its captures the target person's attention alone from multiple persons and it should not interrupts the attention of others. Looking

978-1-4799-5180-2/14/$31.00 ©2014 IEEE

toward a target person by turning gaze/face communicates attending at him/her. If the target human does not respond, s/he will attempt to gain their attention with strong signals (e.g., waving a hand, shaking one's head, moving one's body, or even raising one's voice). In a natural HRI scenario, robots should utilize the same conventions as humans. In other words, the robot should start with a weak action to avoid attracting the attention of individuals other than the target person, then use stronger actions if its initial attempt fails to attract his/her attention.

- **Making eye contact**: Eye contact is a phenomenon that occurs when two people cross their gaze. Gaze crossing function alone is not a sufficient function to make an effective eye contact with other [22]. Gaze awareness in response to other looking response is also an important function to convey someone's intention explicitly for making eye contact. Gaze awareness behavior of the robot will helps to create the feeling of being looked at of the target human. Thus, to be most effective the robot should also display gaze awareness explicitly through some actions (e.g., facial expressions or nodding) after gaze crossing.
- **Shifting attention**: Attention shift involves an agent gazing at or turning to the object referred by another agent [26]. Speakers can direct their partner's attention to objects by exploiting any body parts which they can create an connection [27]. That connection must focus the addressee's attention on the object. Humans are exceptionally good at inferring where others are looking [28] and quite accurate at detecting one another's face-directed gaze at normal conversation distances.

IV. System Overview

For HRI experiments, we developed a robotic head as shown in Fig. 3. This figure shows a prototype of the robotic head and the corresponding outputs of its several software modules. The head consists of a spherical 3D mask, an LED projector (3M pocket projector, MPro150), a laser range sensor (URG-04LX by Hokuyo Electric Machinery), three USB cameras (Logicool Inc., Qcam), and a pan-tilt unit (Directed Perception Inc., PTU-D46). The LED projector projects CG-generated eyes on the mask.

The system utilizes two general purpose computers (Windows XP), each connected to a USB camera. One PC is installed with the head detection and tracking modules (HDTM) for tracking the head of the non-target participant. To detect, track, and compute the direction of the participant's head in real time, we use FaceAPI [29] by Seeing Machines Inc. It can measure 3D head direction within a 3^0 error at 30 frame/second. The other PC is installed with three software modules, namely HDTM, the body tracking module (BTM), and the robot control module (RCM). This PC detects and tracks the bodies of two people in the range sensor data using the BTM. A detailed description of the BTM is provided in [30]. Both PC's are connected through a wired network. The RCM consists of four sub modules: a

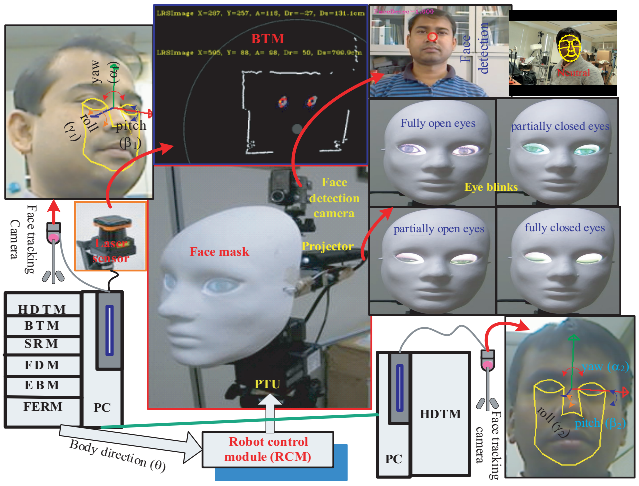

Fig. 2. System overview: consisting of several software modules including HDTM, BTM, SRM, FDM, EBM, FERM, and RCM.

situation recognition module (SRM), face-detection module (FDM), eye-blinking module (EBM), and facial expression recognition module (FERM). The RCM integrates all sensor processing results of both computers.

To assess the current situation (i.e., where the human is currently looking), one observes the head direction estimated by the HDTM. From the results of the HDTM, the SRM recognizes the existing viewing situation (with 99.4% accuracy) and the direction of the relevant object in terms of yaw (α), and pitch (β) movements of the head by using a set of predefined rules. The details description of each rule are found in [31]. The results of the SRM are sent to the PTUCM to initiate the attention control process, and the robot turns toward the human based on the results provided by the BTM. The robot considers that the participant has responded successfully to its actions if the TA looks at the robot within the expected time-frame. If this step is successful, the FDM uses the image from the forehead camera to detect the front of the target participant's face. We use the face detector, which consists of cascaded classifiers based on AdaBoost and Haar-like features [32]. The system recognize the facial expression of the participant by using FERM [33]. The FERM send to the results to EBM for exhibiting eye blinks to let the human know that the robot is aware of his/her gaze. The EBM generates the eye blinks at a rate of 1 blink/second up to three times. All the robot's head actions are performed by the PTU, with the actual control signal coming from several modules.

A. Behaviors Protocol

An shifting attention event is executed by a finite-state-machine model as shown in Fig. 3. To initiate the shifting attention process, the robot (R) begins to observe the current direction of the human's attention by tracking his/her head and body. After recognizing the viewing situation of the target human (TH), the R employ its action plan according to the recognized situation. The R uses the head turning (HT) action, if it recognized the TH in its CFOV situation. It applies head shaking (HS) action for PFOV , and uttering reference terms (URT) for OFOV situations respectively. However, the robot waits about 2 seconds after each attempt for the TH to respond by looking in its direction. Psychological studies shown that 2 seconds delay is enough to capture the human's attention by the robot [22]. For the HT action, we adjusted the pan speed of the pan-tilt unit at $120^0/second$. For the head shaking action, the robot shook its head back and forth ($\pm 30^0$) from its initial position. This meant that the robot turned its head 30^0 left and 30^0 right. The head-shaking speed was adjusted at $240^0/second$. In order to produce URT action, we use a recoded voice ('excuse me').

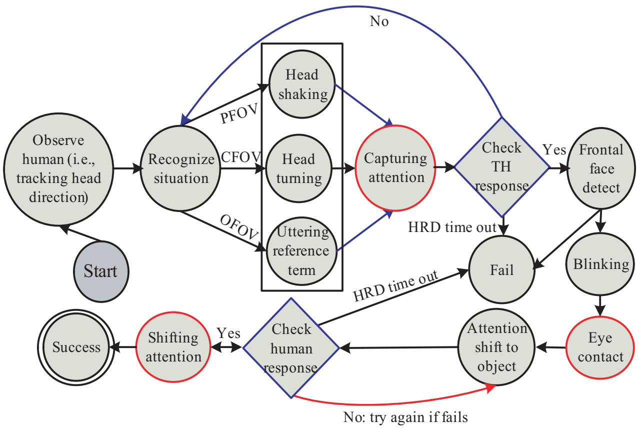

Fig. 3. System overview: consisting of several software modules including HDTM, BTM, SRM, FDM, EBM, PTUCM, and RCM.

The robot considers the TH have responded to its actions if he/she looks at the robot within the expected time frame. It is able to recognize whether this is so by detecting the front of his/her face in the camera image. The robot performs a blinking action to display gaze awareness. The robot avoids the attention of non-target agents (NTH) by turning its head away by -120^0 at a pan speed of $180^0/second$ in the event that an NTH looks at the robot before the TH does.

The robot can be judged to have successfully acquired control of the TH's attention when the latter is gazing at or turning toward the object referred to by the robot. Gaze cues, when combined with head cues, can use effectively to signal attention shift request to the partner. Thus, in our design, we have implemented the shifting attention component of the robot in the form of both eyes and head turns. That means, R makes an attention shift request by switching its attention through turning both eyes and head to the intended object. After addressing the focus based on each cue, the robot waits for up to 2 seconds and checks whether or not the attention shift attempt has been successful. In a successful outcome, both the TH and the robot are looking at the same object, prompting the robot to generate a beep to indicate attention control success. Otherwise, the robot will try again by shifting its eyes and head back toward the TH, and once again focus on the object. The robot considers the case to be a failure if it cannot shift the TH's attention within two attempts.

V. An HRI Experiment

We conducted an experiment to verify that our proposed approach is useful for shifting the target human attention from one direction to the robot's intended direction. 20 graduate students (14 males, 6 females, average age 26.8) at Saitama University participated in the experiment.

A. *Design and Procedure*

In order to simulate a low attention-absorption task, we considered a scenario: 'watching the paintings'. To To prompt participants to look in various directions, we hung seven paintings (P1-P7) on the wall at the same height (just above the eye level of the participants). These paintings were placed in such a way that, when observed from a participant's sitting position, they covered their whole field of view (close to 180^0). To produce the stimuli, we prepared two robotic heads with the same appearance. One robot is called the static robot (S-robot) and it was stationary at all times. The other is called moving robot (M-robot) and it performed the shifting attention process. The placement of robots was exchanged randomly. Two objects were placed on the table as the shifting attention targets.

A pair of participants were asked to sit down on chairs and to look around at the paintings. Since the positions of the robots were fixed, the participants detected them in their various fields of view as they moved their heads and bodies around. The participants took part in six trials in each condition. The moving robot randomly chose one of the participants as the target and the other as the non-target. Each pair of participants interact in three session (each session corresponds to each viewing direction)and lasting approximate 120 seconds. Fig. 4 shows scenes of the experiment.

The robot tried to capture the TH attention while he/she was looking at different paintings so that it could obtain data for three types of viewing situations: the central field of view [e.g., looking at pictures P1 or P2], the near peripheral field of view [e.g., looking at pictures P3, P4 or P5], and the far peripheral field of view [e.g., looking at pictures P6 or P7]. The M-robot initiates the shifting attention process after recognizing the TH's viewing situation. Two video cameras were placed in appropriate positions to capture all interactions during the experiment.

B. *Conditions*

We proposed a robot that seeks to shift a target human's attention with gaze awareness (eye blinks) functions. It is important to compare the proposed robot with another robot the lacks eye blinks functions in order to assess how these functions affect the performance of an interactive task. Moreover, it is also necessary to evaluate the effectiveness of the shifting attention function by comparing the proposed robot with combined eye and head turning action to a robot that only eye turning action. Thus, the proposed model was compared with following two alternative methods. However, the robot performed attention attraction actions behaviors as the proposed for all groups of participants.

(a) The M-robot and the TH making eye contact

(b) The M-robot and the TH are looking at the left object

Fig. 4. Human-robot interaction scenes.

(i) **Method 1 (WEB+ET+HT)**: The robot did not blink its eyes and would stay at its current position for about 3 seconds when the target human (TH) looked at it. After gaze crossing, the robot turn both its eyes and head toward the target object as an attention shift behavior. The robot avert its head from its current position when the non-target Human (NTH) looked at it before the TH.
(ii) **Method 2 (EB+ET)**: The robot blink its eyes about 3 seconds when the target human (TH) looked at it. After gaze crossing, the robot turn its eyes alone toward the target object as an attention shift behavior. The robot avert its head from its current position when the non-target Human (NTH) looked at it before the TH.
(ii) **Method 3 (EB+ET+HT)**:This uses our proposed method described in Section IV-A.

C. Measurements

Quantitative measures: For overall evaluation, we measured the following items by observing the experimental videos.

- **Success ratio (SR)**: This was measured by the ratio of the number of TH looking at the object and the number of times that the robot attempted to shift their attention to that object.
- **Time (T)**: This is the total time required to shift the TH's attention that is indicated by the robot's attention shift cue. This time is measure in terms of time elapsed from when the attention shift actions were begun until the TH started to move his/her attention (i.e., their gaze and/or head) to attend to the target object following the robot's attention shifting cues.

TABLE I
QUANTITIVE MEASURES. VALUES INDICATES THE MEAN AND STANDARD DEVIATION.

	Conditions		
Measurements	Method1 WEB+ET+HT	Method2 EB+ET	Method3 EB+ET+HT
SR	0.60 (0.5)	0.74 (0.45)	0.94 (0.25)
T (seconds)	3.01 (0.27)	2.55 (0.31)	1.72 (0.27)

Subjective measures: After interacting with the all methods, we asked participants to fill out a questionnaire. The measurement was a simple rating on a Likert scale of 1 to 7 where 7 is the highest. The questionnaire contained two items.

- **Impression of shifting attention**: Did you think that the robot was successful in shifting your attention toward the direction it intended?
- **Impression of communicative intention**: Did you think that the robot had any intention to communicate with you?

D. Results

The experiment had a within-participant design, and the order of all experimental trials was counterbalanced. We observed a total of 90 (3 [methods] $\times$ 10 [target human] $\times 3$ [session]) interactions of the target humans for all conditions.

1) Quantitative Evaluation: Table I shows the results of the quantitative measurements.

We conducted a repeated measure of analysis of variance (AVOVA) on the success ratios measurement that showed significant differences among the conditions $(F(2,58) = 4.7, p < 0.001, \eta^2 = 0.10)$. Multiple comparisons with the Bonferroni method showed that the proposed method was significantly better than the methods 1 (p¡0.002) and 2 (p¡0.49) in shifting the target human's attention. No significant difference was found between methods 1 and 2 $(p = 0.29)$ In concerning the time, repeated measure of aalysis of variance (AVOVA) also revealed a significant differences among the conditions $(F(2,18) = 45.07, p <= 0.01, \eta^2 = 0.79)$. Multiple comparisons with the Bonferroni method showed significant differences between pairs (proposed method vs. method 2: $p < 0.002$ and proposed method vs. method 1:$p < 0.01$).

2) Subjective Evaluation: Fig. 5 shows the results of the questionnaire assessment. We conducted a repeated-measure of analysis of variance (ANOVA) on the target participant scores. The results show that the differences among conditions were statistically significant for impression of shifting attention $(F(2,18) = 17.42, p < 0.001, \eta^2 = 0.58)$. Multiple comparisons with the Bonferroni method show significant differences between pairs (method 3 vs. method 1: $p < 0.003$, and method 3 vs. method 2: $p < 0.005$). However, no significant difference was found between method 1 and 2 $(p = 0.75)$ Concerning the impression of communicative intention, a significant main effect was found $(F(2,18) = 101.42, p <$

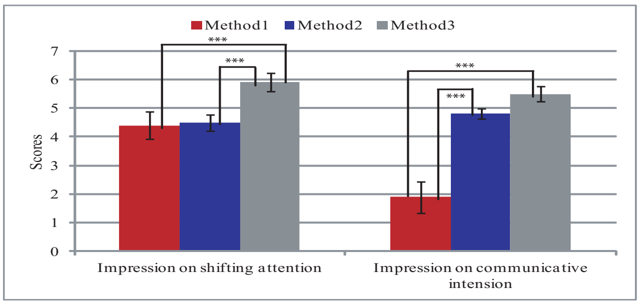

Fig. 5. Result of the post-questionnaire analysis. Errors bars indicate the values of standard errors and *** means significant differences.

$0.01, \eta^2 = 0.89$). Here too the Bonferroni method comparisons revealed that the participant rated method 3 significant higher than that of the method 1 ($p < 0.01$) and method 2 ($p < 0.009$) respectively.

Results revealed that the shifting attention performance of the proposed robot is clearly more effective, in terms of producing a higher success ratio and requires less amount of time, when it employs blinking action in eye contact phase and both eyes and head turning actions in shifting attention phase. A participant's eyes coupled with the robot's eyes during blinking, and this explained why the human participant could quickly identify the robot's shift in attention focus. However, without blinking the robot may fail to create the feeling of eye contact being established, due to its lack of a gaze awareness function. Moreover, the robot was successful at conveying its communicative intention to the target participants using the eyes and head shift functions.

VI. Conclusion

The main objective of our work is to develop a robotic framework that can shift a particular person's attention from one direction to a particular direction by multi-party setting. For this purpose, we have proposed an integrated approach of shifting attention that consists of three phases. Although there may be various behaviors, we incorporated head turning and shaking, eye blinking, and shifting the eyes and the head in respective phases. We have shown that the proposed method can control a target participant's attention in terms of initially capturing his/her attention, making eye contact, and finally shifting his/her attention in the intended direction. Current work consider a two-participant in a controlled laboratory environment. In future, we would like consider the crowded situations and more attention demanding tasks.

References

[1] C. Kleinke, "Gaze and eye contact: A research review", Psychological Bulletin", vol. 100, no. 1, pp. 78-100, 1985.
[2] T. Farroni, E. M. Mansfield, C. Lai, and M. H. Johnson, "Infant perceiving and acting on the eyes: Tests of an evolutionary hypothesis", J. of Exp. Child Psy., vol. 85, pp. 199-212, 2003.
[3] J. Hoffman, Visual Attention and Eye Movements, Psychology Press Ltd, London, 1988.
[4] F. Kaplan, and V. V. Hafner, The challenges of joint attention in infancy, Euro. J. of NeuroSci., vol. 7, no. 2, pp. 135-169, 2006.
[5] M. Staudte, and M. W. Crocker, Investigating joint attention mechanisms through spoken human-robot interaction. J. of Cognit., vol. 120, no. 2, pp. 268-291, 2011.
[6] Y. Nagai et al., A constructive model for the development of joint attention, J. Conn. Sci., vol. 15, pp. 211-229, 2003.
[7] E. Carlson, and J. Triesch, "A Computational model of the emergence of gaze following", Con. Mod. of Cog. & Perc., pp. 105-114, 2003.
[8] G. Heidemann, R. Rae, H. Bekel, and H. Ritter, "Integrating context-free and context-dependant attentional mechanisms for gestural object reference", J. of Mac. Vis. & App., vol. 16, pp. 64-73, 2004.
[9] B. Schauerte, J. Richarz, and G. A. Fink, "Sailency based identification and recognition of pointed at objects", In Int. Conf. on IROS, pp. 4638-4643, 2010.
[10] Y. Nagi, "Learning to comprehend deictic gestures in robots and human infants", In: Int. Work. on ROMAN, pp. 217-222, 2005.
[11] H. Sumioka, K. Hosoda, Y. Yoshikawa, and M. Asada, "Acquisition of joint attention through natural interaction utilizing motion cues", J. of Adv. Robo., vol. 21, no. 9, pp. 983-999, 2007.
[12] H. Kozima, and A. Ito, "Toward language acquisition by an attention sharing robot", In Joint Conf. on New Methods in Lang. Proc. & Comp. Nat. Lang. Learn., pp. 245-246, 1998.
[13] N. Kubota, T. Shimizu, and M. Abe, "Joint attention between a human being and a partner robot based on computational intelligence", J. of Adv. Comput. Intelli. & Inte. Info., vol. 11, pp. 1274-1280, 2007.
[14] L. F. Marin-Urias, A. Sisbot, K. Pandey, and R. Alami, "Towards shared attention through geometric reasoning for human robot interaction", In: Int. Conf. on Humanoid Robots, pp. 331-336, 2009.
[15] Z. Yucel, A. A. Salah, C. Mericli, and T. Merili, "Joint visual attention modeling for naturally interacting robotic agents", In Int. Symp. on Com. and Info. Sci., pp. 242-247, 2009.
[16] H. P. Huang, J. H. Chen, and H. J. Jian, "Development of the joint attention with a new face tracking method for multiple people", In IEEE Work. on Adv. Robo. & Its Social Imp., pp. 1-6, 2008.
[17] Z. M. Hanafiah, C. Yamazaki, A. Nakamura, and Y. Kuno, "Understanding inexplicit utterances using vision for helper robots", In: ICPR, pp. 925-928, 2004.
[18] O. Sugiyama et al., "Human-like conversation with gestures and verbal cues based on three-layer attention drawing model", J. Conn. Sci., vol. 18, pp. 379-402, 2006.
[19] C. Peters, "A perceptually-based theory of mind for agent interaction initiation", Int. J. Human. Robo., vol. 3, no. 3, pp. 321340, 2006.
[20] C. M. Huang, and A. L. Thomaz, "Joint attention in human-robot interaction", In Dialog with Robots: AAAI, pp. 32-37, 2010.
[21] B. Mutlu, F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita, "Nonverbal leakage in robots: Communication of intensions through seemingly unintentional behavior", In Int. Conf. on HRI, pp. 69-76, 2009.
[22] M. M. Hoque, T. Onoki, Y. Kobayashi, and Y. Kuno, "Effect of robot's gaze behaviors for attracting and controlling human attention, J. of Adv. Robo., vol. 27, no. 11, pp. 813-829, 2013.
[23] M. M. Hoque, T. Onuki, Y. Kobayashi, and Y. Kuno, "Attracting and controlling human attention through robot's behaviors suited to the situation", In: Int. Con. on HRI, pp. 45-46, 2012.
[24] Attention: Definition of attention, http://www.en.wikipedia.org/wiki/Attention, May 2012.
[25] S. R. H. Langton, R. J. Watt, and V. Bruce, Do the eyes have it? cues to the direction of social attention, Tre. in Cog. Sci., vol. 4, no. 2, pp. 50-59, 2000.
[26] G. Butterworth, "The ontogeny and phylogeny of joint visual attention", in Natural Theories of Mind: Evolution, Development, & Simulation of Everyday Mindreading, pp. 223-232, 1991.
[27] H. H. Clark, Using Languge, Cambridge: Cambridge University Press, 1996.
[28] F. Delaunay, J. D. Greeff, and T. Belpaeme, "A study of a retro-projected robotic face and its effectiveness for gaze reading by humans", In: Int. Conf. on HRI, pp. 34-44, 2010.
[29] FaceAPI: Face Tracking for OEM Product Development, Seeing Machines Int., http://www.faceapi.com, 2010.
[30] Y. Kobayashi, Y. Kuno, "People tracking using integrated sensors for human robot interaction", In: Int. Conf. on IT, pp. 1597-1602, 2010.
[31] M. M. Hoque, D. Das, Y. Kobayashi, Y. Kuno, and K. Deb, "Proactive Approach of Making Eye Contact with the Target Human in Multi-Party Settings", In: Int. Conf. on ICCIT, pp. 191-196 , 2013.
[32] G. Bradsky, A. Kaehler, V. Pisarevskky, Learning based computer vision with Intel's open source computer vision library, J. Intel Tech., vol. 9, pp. 119-130, 2005.
[33] M. H. Mahoor, M. A. Mottaleb, and A. N. Ansari, "Improved active shape model for facial feature extraction in color images", J. of Multimedia, vol. 1, no. 4, pp. 21-28, 2006.