2019International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-9 February, 2019

Developing a Bidiretional Mutual Gaze Mechanism for Human Robot Interaction

Shayla Sharmin
Dept. of Computer Science and Engineering
Chittagong University of Engineering and Technology
Chittagong, Bangladesh
e-mail: shayla.turin@gmail.com

Mohammed Moshiul Hoque
Dept. of Computer Science and Engineering
Chittagong University of Engineering and Technology
Chittagong, Bangladesh
e-mail: mmoshiulh@gmail.com

Abstract—Establishing mutual gaze is one of the most fundamental capabilities to initiate any interaction in both human-human and human-robot communications. This capability of mutual gaze is particularly important when people collaborate over a shared visual space. This paper proposes a bidirectional mutual gaze mechanism for human-robot communication. It is assumed that look at each other makes the mutual gaze. However, simply looking at each other alone does not ensure mutual gaze. Both agents also need to be aware of being watched by the other. In order to perform mutual gaze effectively, we propose two functions: gaze crossing and gaze awareness. For gaze crossing, the system detects the human face and his/her eye region. The system also tracks the eyeball of the target human by locating the position of eyeball. After gaze crossing, the system detects the smile of human which ensure the gaze awareness. Continuing the flow of communication, the robot performs eye blinking, head nodding and verbal actions in response to human. Several software modules are developed to perform the bidirectional mutual gaze mechanism and evaluate the effectiveness of each module with several experimental trials. Moreover, the system is tested in human initiative and robot initiative mutual gaze cases.

Keywords— Human robot interaction; bidirectional mutual gaze; face detection; gaze tracking; gaze awareness.

I. INTRODUCTION

Robots are rapidly being developed for real world application areas to interact human and so it is an obvious need to have some intelligence in the robots for a friendly interaction with their daily activities. The basic goal of HRI is to develop human like interface and naturally interact with human. A number of verbal and non-verbal mechanisms such as body movement, facial expressions, gesture and posture are used when people interact. Controlling the flow of communication is the main role of gaze in human communication system [1] as it is acknowledged that eyes can be used for both giving signal and perceiving the state. For human beings, this dual function makes the eyes a remarkable tool for social interaction. The bidirectional gaze mechanism can improve coordination in interaction by correcting potential failures before they occur in a subtle way [2].

Under these above circumstances, we can be in a conclusion that to make a friendly human-like environment a robot should have the ability of bidirectional mutual gaze mechanism by detecting the interest of human along with the ability to take steps to start the conversion. In this paper, we have developed a bidirectional mutual gaze mechanism which helps to start interaction between human and robot that consists of two types of event. In the first case when the human wants to communicate with the system he/she looks at it and smile projecting cordial invitation and after detecting his/her interest the system starts communication. On the other hand, considering the fact when a robot, for example, the clerk robot, the guiding robot, the home service provider robot or a seller robot in a shopping mall, wants to communicate with the human our proposed system should first catch any human being whom it desires to start communication and then fix its gaze on that human. By considering both cases, we hypotheses that our system performs two tasks consecutively for making eye contact with the human: (i) Gaze crossing that is to look each other eyes or face, and (ii) Gaze awareness that is displaying an explicit behavior to respond. And to complete these tasks, face, eye and smile has been detected using Haar features before starting exchange.

II. RELATED WORK

When two persons look at each other which define crossing their gaze, it is considered that mutual gaze phenomenon establishes which plays an important role in initiating iteration and in regulating face-to-face communication, social communication and perception of others' emotion states. Psychological studies demonstrate that the gaze crossing act alone may not be sufficient to create a successful eye contact event [3, 4]. So, we need gaze-awareness action to establish a positive conjoint gaze which is the significant deeds for individuals to sense that the robot recognizes his/her attention reaction [5]. To express awareness, the robot ought to use certain actions that can be verbal or non-verbal (i.e. shout, blink eyes or smiling or both). The Psychologist says that mutual gaze enrich the feeling of attraction, excitement, surprise, love, faith, fascination, invitation, gravitation with each other. The main role of gaze in human communication is to control the flow of communication which is called meta-communication and recently several robots such as Robita [6], Robovie [7] and Cog [8] are utilizing gaze for meta-communication.

Eye Contact Robot first zoomed in the image and then detects the eyes (pupils) as well as the nostrils using a flat screen to display their graphical head which is unnatural as a face [2]. Robita are invented to make eye contact with human by revolving their eyes toward human faces that is not enough to continue further communication [9]. Robovie R-2 robot was developed in a more effective way to actively participate in a gaze tracking with the human participants, but that the goal of this project could not be achieved [10]. A system that is so far an extensive application of HRI they use faceAPI to detect the human face [5]. The main concern of these robots is costly & complicated to design, maintain & construct.

III. MODELLING BIDRIECTIONAL MUTUAL GAZE

A. Basic Eye Contact Method

Before starting communication, it is necessary to fulfill two conditions which are gaze crossing and gaze awareness to feel that there is a successful eye contact. It is stated that in case of bidirectional communication, both parties can initiate conversation by mutual gaze. Generally in two parties' communication, one participant provides instruction to other participant and then further action is taken.

Gaze means to look steadily, intently, and with fixed attention. Most human research literature on gaze is concerned with the direction of eyes while eyes are the key source of information about the direction of responsiveness, social gaze involves a complex coordination of the eyes the head and body orientation. Mutual gaze occurs when both parties look towards each other's face or eye region thus acting simultaneously as sender and recipient. Mutual gaze is a phenomenon that establishes when two people cross their gazes which plays an important role in initiating iteration and in regulating face-to-face communication, social communication and perception of others' emotion states. After crossing their gazes, both parties need to be aware that they are looking at each other and willing to start communication which refers as gaze awareness. Fig. 1(a) shows gaze crossing which is the first condition of mutual gaze whereas Fig. 1(b) demonstrates that they both are aware to further communication by raising gaze awareness (i.e. smiling) that is displaying an explicit behavior to respond.

(a) Gaze Crossing (b) Gaze Awareness

Fig. 1. Bidirectional Mutual Gaze

It is obvious that if someone turns his or her gaze to a target person with a positive facial expression the chance of avoidance from the other person is less than with a negative or no facial expression [11]. Blinking eyes and nodding head are also good entities for gaze awareness, but the smiling face conveys the message of affirmative movement without any misunderstanding.

B. Conceptual Model of Human Robot Interaction

To design a robotic system, we should make sure that the response time of a robot cannot be so too fast or too slow. The conceptual models of human initiative and robot initiative have been demonstrated in Fig. 2 and Fig. 3. In both case R detects H's face, eyes and smile along with tracking the gaze. Adopting the idea of this "2 seconds rule", [12] reveals that most of the participants of their experiments responded within 2 seconds after the robot takes action to attract attention.

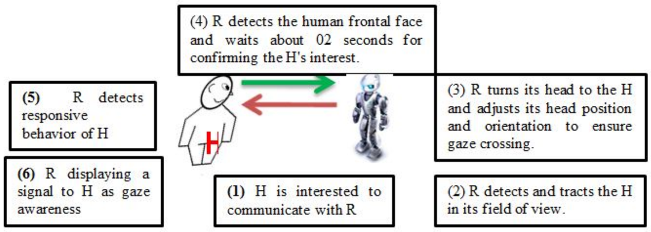


Fig. 2. Conceptual Model for Human Initiative Case

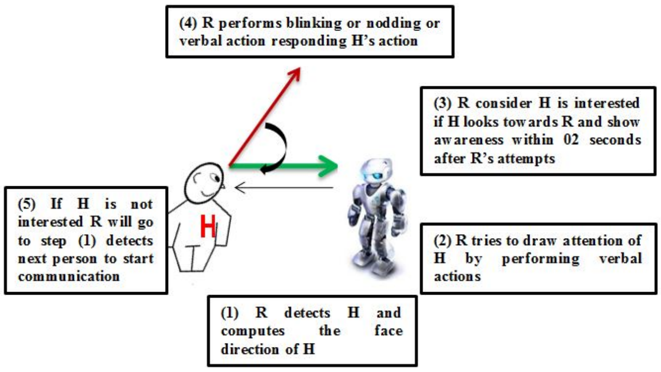


Fig. 3. Conceptual Model of Robot Initiative Case

C. Behavioral Protocol

The core objective of our effort was to develop a mechanism for social robots by which robots can control the target human attention mainly by verbal action. The robot takes the following steps for the following two cases:

1) Case I: Human initiative case:

However, we assume that the human and the robot are now watching in their respective direction that is observing something. If the human is looking at the robot, it takes the following steps:

Step (1): R detects and tracts the H in its field of view.
Step (2): R ensures gaze crossing with H.
Step (3): R detects the human frontal face and waits about 2 seconds for confirming the H's interest.
Step (4): Establish eye contact. R calculates the awareness of H by checking if he/she is smiling. H feels that s/he has made eye contact with R.
Step (5): R performs eye blink or head nod or verbal action to show affirmative feedback in respond to smile of H. Human senses that s/he has ended eye contact with R. Communication starts.

2) Case II: Robot initiative case:

In this case we assume that the robot wants to start communication with the human. And so the robot first finds the person whom it desires to start and takes the following steps:

Step (1): R searches for H.
Step (2): R stops searching after detecting human's head.
Step (3): R waits (2 sec) for H's gaze. If H gazes at R, they are now in face to face and go to step 5. Otherwise, R performs verbal action to control human attention.
Step (4): R performs the first attempt to draw attention of human and waits for 2 sec, if H looks at R go to step 5 otherwise take second attempts to control the attention and waits for 2 sec. If a positive result then takes step 5 otherwise, R gives up controlling the human attention.

Step (5): R detects H's face. R considers that H has responded to its actions if H looks at R within expected times and smile towards it, which is recognized by detecting H's frontal face in the camera image.
Step (6): R performs eye blink or head nod or verbal action to show gaze awareness. H feels that s/he has made mutual gaze with R.

D. Hardware Setup

We have developed a robotic platform for our HRI experiments shown in Fig. 4. The robot head rotates with the help of servo motors which is controlled by Arduino uno. Two servo motors have been used; one for tilt movement and another is for pan movement.

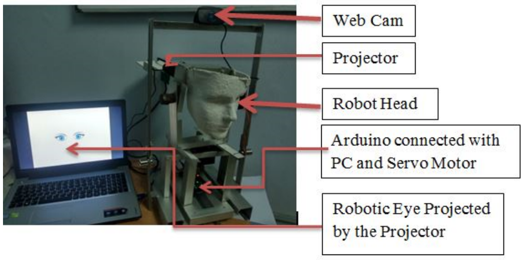


Fig. 4. Hardware Setup

The eye of the robot is projected by the 3M MPro Pocket Projector on the tracing paper attached on the eye portion of the robot head. The web cam is attached to the above of the robot head. The servo motor rotates from left to right and if catch a human face the signal is sent to the Arduino and it stops rotating.

E. Software Configuration

The system used one general purpose computer connected with a USB camera. The robot turns its head to find a target human to communicate. It stops moving its head when it finds a human in its level of view. The result of the face detection module (FDM) is sent to the robot control module (RCM). To compute the participant's situation viola jones algorithm [13] has been used to detect face, eyes and smile. If a person is found the result is sent to the RCM, then robot stop moving. The gaze tracking module (GTM) tracks the eyeball if it is looking straight towards the robot or not. If H is not looking at R, R performs verbal action to draw attention that the Gaze awareness detection module (GADM) detects if the H is smiling or not. If the result is positive this will consider as gaze awareness establishing the theory that H understands that the R wants to communicate and R response with greetings and mutual gaze establishes. If R fails to establish the communication within two attempts of drawing attention, R gives up trying controlling H. The whole software module is divided into three sections; the face detection module, the gaze tracking module and the gaze awareness module. The detections are performed by viola jones algorithm.

1) Head Detection Module:

This module can be divided into two parts, one is the side face detection and other is the frontal face detection module. The detection is processed using Haar features. First, R detects human, and then checks if it is in face to face condition or not. A web cam searches for human. The image is sent to the module in gray scale captured by the webcam, and then face is detected using Haar features which are considered as relevant for human frontal or side face [14]. The Haar cascade returns the coordinates (x,y) of the upper left corner values of the face region in the given image with the height (h) and width (w) where it detects a face. Fig. 5 shows a blue rectangle around the face.

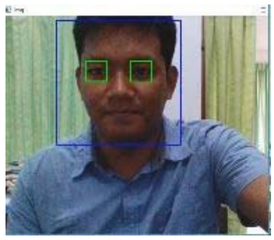


Fig. 5. Detecting Face and Eye

*2) Gaze Tracking Module:*The gaze tracking module checks if the eye is looking straight towards R.People look at a certain point by two ways; the first way is by turning the head and eyes where other is only turning the eyes towards a specific direction. Turning both eyes and head is considered as the highest level of priority to communicate to other. Only finding frontal face cannot decide whether the person is looking at the R or not. In many cases it is possible to find a situation that R finds a frontal face but the H is looking at the other direction. To overcome this problem we have developed a module which detects the eye gaze and if the eye is looking straight to the R, the result is delivered to the system considering establishment of the gaze cross. In Fig. 5 eye portions is selected by the green rectangle. Fig. 6 explains how we track the gaze. The procedure is described briefly in the following sectionshowing the output in Fig. 7.

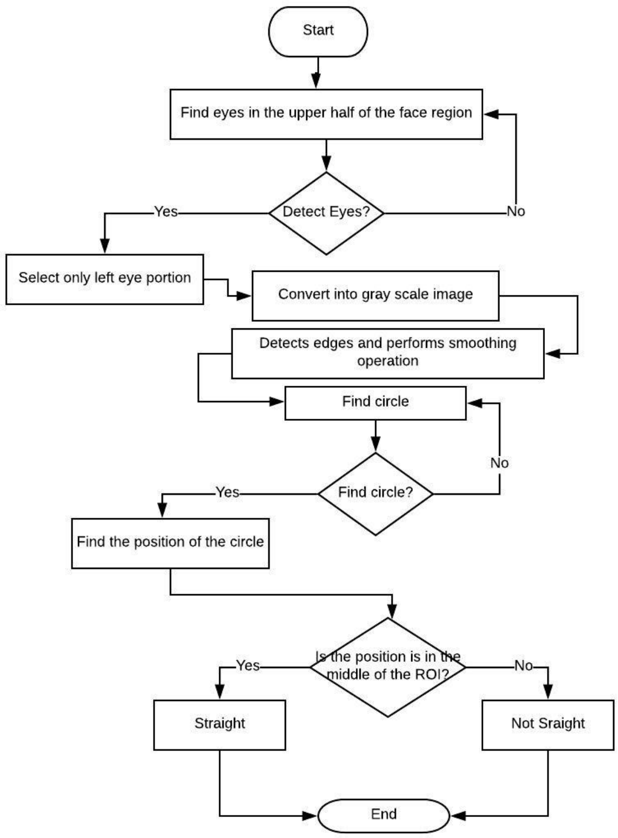


Fig. 6. Flow Chart of GazeTrack Module

(a)Left eye (b)Gray Conversion (c)Edge Detection (d)Smoothing (e)Detect Circle

Fig. 7. Gaze Tracking Module

*a) Detect Eye Region:*From the Detected Face, the region for performing the eye search is created. The region of the upper of the face isisolated for performing the eye search. The eyes are found in the upper half region. This eye region is then isolated by searching the region [y:y+h/2,x:x+w] where y is the x and y are coordinates along with w and h where they are the width and height value of the face region. An eye search in the eye region is performed using similar Viola-Jones[14] based method. And this returns the ex,ey,eh and ew values.

b) Crop portion of one eye: As in normal case. both eyes directed in the same side so working with one eye will be enough to detect the position of the eye (Fig. 7(a)). It is not a matter of choosing right or left eye. This helps us to less the computational procedure. To isolate one eye we crop the area defined by the following equation:

$$one_{eye} = ey{:}\,ey + \frac{eh}{2}, ex{:}\,ex + \frac{ew}{2} \text{.........................(1)}$$

*c) Gray Scale Conversion:*The next step is to convert the eye region into gray scale. We have converted based on Luminance [15] is designed to match human brightness perception by using a weighted combination of the RGB channels based on the following equation shown in Fig. 7(b).

$$g_{Luminance} = 0.3R + 0.59G + 0.11B \text{......................(2)}$$

*d) Detecting Edge:*The next step is to detect the edges of the eye region. We used the canny edge detector [16]to detect the edges shown in Fig. 7(c) as it is considered as the perfect one.

e) Smoothing ROI: This part is performed to smooth the image to reduce noise by performing blurring operationFig. 7(d). Image blurring is done by convolving the image with a low-pass filter kernel which is useful for eliminating high frequency content (i.e. noises, edges) from the image. Edges are blurred slightly in this process and to do so we acquired averaging method which is done by convolving image with a normalized box filter. It basically takes the average of all the pixels under kernel area and substitutes the central elements.

*f) Detect Circle:*The most crucial part is to detect circle in the given image. Circle denotes the iris of the eye, by detecting its position the system will calculate if it is straight or not A circle is represented mathematically as

$$(x - x_{center})^2 + \left(y - y_{center}\right)^2 = r^2 \text{ .... (3)}$$

where (x_{center}, y_{center}) *is the center of the circle, and* r *is the radius of the circle.*

In a two-dimensional space, a circle can be designated by the above equation. We have used 21HT circle detection algorithm discussed in [16]. The circle is drawn in Fig. 7(e) on the image.

g) Calculate the Position of the Eyeball: In this section the center of the eyeball is detected in respect of the eye region. For this, we check the location of the center of the eye calculating by Hough circle detection algorithm. If the eyeball is located at the center a positive result will send to the system as a confirmation of theeye cross establishment.

To detect the position of the eye ball we just simply calculate the distance of the center from the upper (*disU)*, lower *(disD),*left*(disL)* and right*(disD)* boundary from the left eye region.

$$d_i = \sqrt{(x_r - x_i)^2 + (y_r - y_i)^2} \text{ .... (4)}$$

where x_r, $y_r$are the center of the circle and x_i, y_i are coordinates of upper, down, left and right boundary.

If $disU \cong disD$ *and* $disL \cong disR$; we can say that the eyes are looking straight to the robot.

In Fig. 8 we can see that the H is looking towards the screen as we can see that the eyeball is surrounded by a circle where Fig. 9 shows the scenario when H is not looking towards the R.

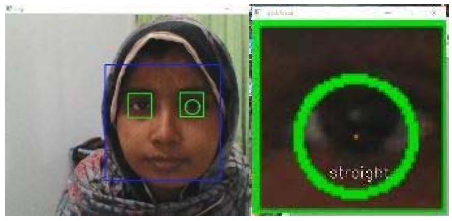


Fig. 8. H looking towards R

Fig. 9. H is looking at another direction

*3) Detecting Awareness:*Grasping the facial expression is much easier to understand the gaze awareness rather than eye blinking which was adopted by previous work.So in this paper smile is reflected as the gaze awareness which is a stronger entity than the eye blinking. In this case it is noticeable that when people laugh there is a change in the eye area and mouth area. But in case of courtesy smile only mouth area is changed, as the jaw is stretched out, but the changes in the eye region is negligible. For the detection purpose Haar cascade is used. Haar cascade returns a Xs, Ys, Ws and Hs, where Xs,Ys are the coordinates of the left upper corner of the smile region and Ws,Hs are the width and height of the returned area shown in Fig. 10.

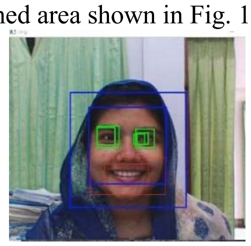


Fig. 10. Detection of Smile

IV. Experimental Setup

To test the modules we have used a laptop and a webcam in the real world environment. Total 10 peoples participated in the experiment. All of them are graduate students of the Department of CSE, Chittagong University of Engineering and Technology. Fig. 11 shows an example scene of the experiment.

Fig. 11. An example scenario from the experiment where a subject is standing in front of the robot.

A. *Human Initiative case:*

In this case R checks if H is looking straight towards the R to ensure gaze cross and then determines if the H is smiling within 2 seconds. That means gaze awareness has been detected and H is interested to start further action and so R replies to H and the scenario is depicted in Fig. 12.

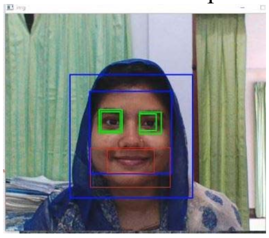
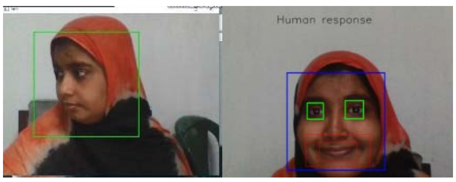


Fig. 12. R finds frontal face; H is looking straight towards R and smiling; Human initiative case established

B. *Robot Initiative Case:*

Fig. 13 shows if in between 2 seconds (after detecting human) there is no gaze cross then R tries to attract human by verbal action and again waits for 2 seconds for the response. If there is no response from H in 2 seconds R tries again. If the H turns his/her head to R and smile that means R's attempt to attract human is successful.



(a)H is Busy *(b) H responds*

Fig. 13. H is busy; R performs verbal actions; H turns head; gaze cross and gaze awareness (smile) is detected; mutual gaze is established

V. Result Analysis

Three software modules are used to ensure functionalities of proposed bidirectional mutual gaze mechanism. The experimental result contains a human face as input from the physical environment and generates the corresponding robotic gaze as the human moves around through the field of view of the robot. For the evaluation of the developed system we performed a series of experiments with different criteria. The data from 10 peoples (03males and 07 females), in different positions and lighting condition were collected and analyzed at the operating system lab in Chittagong University of Engineering and Technology. In case of gaze track we only consider when the eye is open.

To calculate the detection accuracy, we have captured the experimental scenario by screen recording software. We have collected data of the face, eye and smile from these videos with different ranges from 10 peoples. Fig. 14 shows the histogram graphic on the basis of the mean and the standard deviation value of detection rate. The recorded video was ranging 10-20 seconds long and the frame rate per second was approximately 25-31.Total 4055 frames that means in average 405.5 frames are considered in our experiments where human are present. In an average 405.5 frames the mean (SD) of face detection and eye detection of 10 peoples are 392.5(SD≅9.19) and 380(SD≅18.03).And in an average 211.2 frames with smiling face the mean value of detected smile is 185.9 (SD≅17.88).

Fig. 14. Column bar indicates mean values and the error bar indicates standard deviation of face, eye and smile detection modules.

To analyze gaze tracking module, we have recorded videos ranging 8-10 seconds by screen recording software. Form an average 256.6 frames, we get 241.1 (SD≅10.25)successfully detected frames where eye is looking straight towards the robot and in an average 99.8 frames the mean value of the situation where H is looking at other direction is 93.4 (SD≅4.5). In this experiment, we considered the frames where the eye is open and then we track the gaze direction. Fig. 15 is the graphical representation of mean and standard deviation of the output of gaze tracking module where 10 peoples performed.

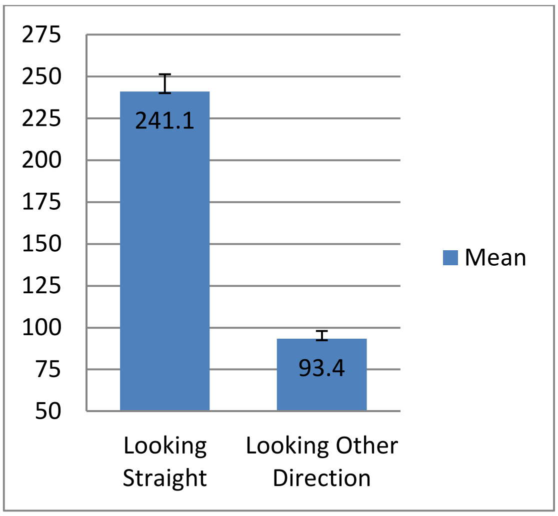


Fig. 15. Column bar indicates mean values and the error bar indicatesstandard deviation of Gaze Tracking Module

Based on these 03 sub modules, we have designed our bidirectional mutual gaze mechanism system. We analyzed both human initiative case and robot initiative cases with 10 persons and each performs 10 trials which means total 100 experiments have taken place. Table I describes that in average 8.4 of 10 trials are successfully recognized by the system in human initiative case and in average 8.2 of 10 trials are perfectly performed by the robot initiative case. And the total success rate of the whole mechanism is 83%. The success rate varies because sometimes the system could not detect the situation of H. In robot initiative case sometimes human respond after 2 seconds and that is the reason the system fail to understand human's intention.

TABLE I. SUCCESS RATE OF PROPOSED MECHANISM

Avg. #of trial	**Avg. # of Successful Attempts**		**Total Success Rate**
	Human Initiative Case	***Robot Initiative Case***	
10	8.4	8.2	83%

VI. CONCLUSION

In the near future robots will play a vital role in our day to day life where it has already made an important place in our social life. In this paper, we have discussed a method which will minimize the gap between human and robot interaction in some extent. After detecting face we focused on the gaze direction which actually helps the situation to avoid any misunderstanding. By selecting smile as the gaze awareness entity, it becomes more acceptable and comprehensible and also the confusion arising rate among human become less because of verbal action and nodding head. For the detection purpose, we have used Haar cascades which sometimes fail to give an accurate result because of different lighting condition and luminance. Gaze tracking module helps the system to detect whether the human is interested or not more accurately. The system faces the problem when there are more than one people and this should be introduced in our project to make it more acceptable in the real life human robot interaction.

ACKNOWLEDGMENT

This work was supported by the DRE project, Chittagong University of Engineering and Technology (CUET).

REFERENCES

[1] D. Miyauchi, A. Nakamura, and Y. Kuno, "Bidirectional eye contact for human robot communication," in *IEICE Tran. on Info. and Sys.,* vol. E88-D. no. 11, pp. 2509-2516, Nov. 2005.

[2] S. Andrist, M. Gleicher and B. Mutlu, "Looking coordinated: bidirectional gaze mechanisms for collaborative interaction with virtual characters", in Proc. CHI, pp. 2571-2582, 2017 .

[3] M. Cranach, "The Role of Orienting Behavior in Human Interaction, Behavior and Environment", in *The Use of Space by Animals and Men*, pp. 217-237, *Plenum Press*, New York, 1971.

[4] Y. Yoshikawa, K. Shinozawa, H. Ishiguro, N. Hagita and T. Miyamoto, "The effects of responsive eye movement and blinking behavior in a communication robot," in Proc. IEEE/RSJ Int. Con. on Intelligent Robots and Systems, Beijing, pp. 4564-4569, 2006.

[5] M. M. Hoque, D. Das, T. Onuki, Y. Kobayashi, and Y. Kuno, "Robotic system controlling target human's attention," in *Intelligent Computing Theories and Applications*, pp. 534–544, Springer, 2012.

[6] G. O. Deák, "When and where do infants follow gaze?," in Proc. *ICDL-EpiRob*, Providence, RI, 2015, pp. 182-187.

[7] A. Maejima, T. Kuratate, B. Pierce, S. Morishima and G. Cheng, "Automatic face replacement for a humanoid robot with 3D face shape display," in Proc. *Humanoids*, Osaka, pp. 469-474, 2012.

[8] T. Kanda, H. Ishiguro, T. Ono, M. Imai and R. Nakatsu, "Development and evaluation of an interactive humanoid robot 'Robovie'," in Proc. *IEEE Int. Con. on Robotics and Automation,* pp. 1848-1855, 2002.

[9] R. A. Brooks, C. Breazeal, M. Marjanovic, B. Scassellati, and M. M. Williamson, "The Cog project: Building a humanoid robot," in *handbook of computation for Metaphores, Analogy and Agents, LNAI*, vol. 1562 ed. C. Nehaniv,pp.52-87, Springer- verlag, 1998.

[10] Y. Yoshikawa, K. Shinozawa, H. Ishiguro, N. Hagita, and T. Miyamoto, "Responsive robot gaze to interaction partner", in Proc. *Robotics: Science and systems*, 2006.

[11] M. M. Hoque, T. Onuki, Y. Kobayashi and Y. Kuno, "Controlling human attention through robot's gaze behaviors," in Proc. *Int. Con. on Human System Interactions, HSI 2011*, Yokohama, pp. 195-202, 2010.

[12] M. L.Pönkänen and J. K. Hietanen "Eye contact with neutral and smiling faces effects on autonomic responses and frontal EEG asymmetry, vol 6, article 122, May 2012.

[13] P. Viola and M. Jones, "Rapid object detection is using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition*, 2001.

[14] W. K.Pratt, Digital image processing. Wiley-Interscience, pp. 61-86, 2007.

[15] J. Canny, "A Computational Approach to Edge Detection," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679-698, Nov. 1986.

[16] H. KYuen, J. Princen, J. Illingworth and J.Kittler, "Comparative study of Hough Transform Methods for Circle Finding" in *Image Vision Computing,* vol. 8, no 1, pp. 71-77, 1990.