BECKONING ROBOTS WITH THE EYES

Y Kuno, D Miyauchi and A Nakamura

Saitama University, Japan

ABSTRACT

Eye contact is an effective means of controlling human communication, such as in starting communication. It seems that we can make eye contact if we simply look at each other. However, this alone does not establish eye contact. Both parties also need to be aware of being watched by the other. We propose a method of eye contact considering these two conditions for human-robot communication. When a human wants to start communication with a robot, he/she watches the robot. If it finds a human looking at it, the robot turns to him/her, changing its facial expressions to let him/her know its awareness of his/her gaze. This completes eye contact. We then present a robot with a wide field of view using three cameras. If a person wants the robot to come to him/her, he can do so just by making eye contact with the robot.

INTRODUCTION

Gaze plays an important role in human communication. Thus, there has been a great deal of research on using gaze or eye movements for human interfaces, which can be considered as communication between man and machine. Most human interfaces use eye movements for pointing instead of a mouse or a joystick, such as in choosing an icon [Hansen et al (5), Jacob (7), Ohno and Mukawa (12)]. However, unlike arms and hands, eyes are not appropriate for precise pointing. We move our eyes toward the direction that objects we would like to see, exist in. However, we do not use our eyes to point something. The main role of gaze in actual human communication is to control the flow of communication. This function can be called meta-communication. Recently, several robots have been proposed that utilize gaze for meta-communication. ROBITA [Matsusaka et al (10)] turns to the specific person speaking at the moment in a group conversation. Robovie [Kanda et al (8)] and Cog [Brooks et al (2)] are similar examples.

Eye contact is a phenomenon that occurs when two people cross their gaze. Since we perceive eye contact clearly, eye contact has stronger meta-communication capability than a simple gaze. Suppose we would like to make an order to a waiter in a restaurant. We search for a waiter, then waiting until he turns toward our direction. When he turns, our eyes meet his eyes. We make eye contact. Usually, this eye contact is enough to let the waiter know that we want him to come over.

The robots mentioned above such as ROBITA are supposed to make eye contact with humans by turning their eyes (cameras) toward human faces. Psychological studies show, however, that this turning action alone may not be enough to make eye contact. In addition, each party must be aware of being looked at by the other [Cranach (3)]. Thus, we have proposed a method of eye contact between human and robot considering the above through the observations of the human face and the actions of the robot head and face [Miyauchi et al (11)]. Since the robot's actions including changes in its facial expression play an important role, we call our method active eye contact.

In our previous work (11), we combined eye contact and gesture for human-robot communication. We have developed a robot that accepts hand-gesture commands only after making eye contact with the person. However, in various cases such as the waiter case mentioned above, we may not need hand gestures. Eye contact alone is often enough to convey our intention. Thus, we modify the robot so that we can control it only through eye contact. The robot should act faster for smooth communication with humans. We use three cameras to realize a wide field of view. The robot can find a person who is looking at it faster. We show that we can make the robot come by eye contact through experiments.

ROBOT SYSTEM

This section briefly describes our first robot system used for eye-contact experiments before explaining our method of making eye contact between humans and robots. Figure 1 shows our robot. We use the mobile robot Pioneer 2 by ActivMedia. A laptop PC is placed on it so that a 3D CG human head is shown at an appropriate height. A pan-tilt-zoom controllable camera (EVI-D100 by Sony) is set above the PC with a black screen behind it so that it will not attract attention and people can concentrate on looking at the face on the computer display.

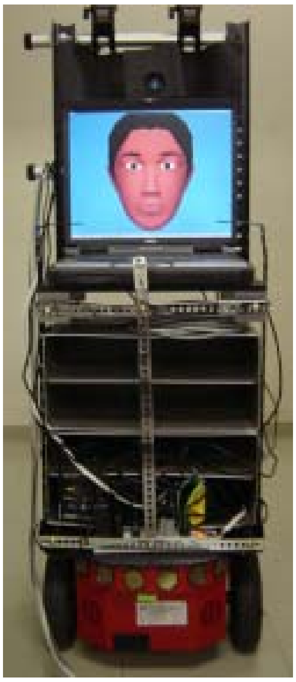


FIGURE 1- Prototype eye-contact robot.

EYE CONTACT BETWEEN HUMANS AND ROBOTS

As mentioned in the Introduction, two conditions – gaze-crossing and gaze-awareness – are necessary for humans to feel that they have made eye contact.

It is relatively easy to satisfy the first condition since this is a sort of physical condition. The robot observes the human's gaze. If the human is looking in the direction of the robot, it turns its eyes toward the human. This observation-and-action sequence can fulfill the first condition.

The second condition is more difficult to satisfy. Even if the robot has noticed that the human is looking at it, the person may not be aware of this fact. We solve this problem by making the robot show this fact explicitly by changing its facial expressions. If the robot finds a human looking at it, it turns its face toward him/her. If the person is still looking at it after this action, the robot assumes that the human is really looking at it. Then, it changes facial expressions, such as by smiling, to let the person know that it is aware of his/her gaze. We again use this observation-and-action sequence to fulfill the second condition.

We have found from preliminary experiments that we need one more thing to realize eye contact between humans and robots. The robot should not make humans feel that it is looking at them when it is actually not. There is, however, the so-called Mona Lisa effect [Kendon (9)]. We humans tend to perceive that a face in a still image is looking at us when we look at it. To avoid this effect, the robot should keep moving its head when any human is not looking at it.

Based on these considerations, we have designed an eye contact method as follows.

1. While rotating its camera, the robot detects face candidates. The CG head on the display turns in the same direction as the camera.

2. If it detects a face candidate, the robot examines the existence of the eyes and the nostrils to confirm whether or not it is a human face. (If not a face, return to 1.)

3. The robot computes the face direction. If the direction is toward the CG head on its display, the robot turns in the direction of the human. It also turns the CG head to the frontal face position. If the robot observes that the human is still facing toward it after this action, the robot changes its facial expressions on the CG head, completing eye contact with the human. (If the face direction is not toward the robot, return to 1.)

FACIAL IMAGE PROCESSING

Psychological studies show that humans may avoid eye contact when they are too close to each other. The frequency of eye contact increases as the distance between the humans increases [Argyle and Ingham (1)]. This means that the robot should be able to make eye contact even when humans are a little far from the robot. Thus, our robot first searches for face candidates with the zoomed-out camera. When a candidate is detected, the camera zooms in on it. Then, the robot examines detailed face features.

Face Candidate Detection

Face candidate regions are detected in the images with a wide field of view (Figure 2). First, skin color regions are extracted (Figure 3). Then, small regions and too elongated regions are removed. Inside the remaining regions, subtraction between consecutive frames is computed. The largest region among those where the sum of absolute values of the subtraction exceeds a given threshold is considered as a face candidate (Figure 4). Then, the pan, tilt, and zoom of the camera are adjusted so that the candidate region can be taken large enough to examine facial features. Experiments show that it can detect human faces indoors at a distance of 6 m.

FIGURE 2- Input image.

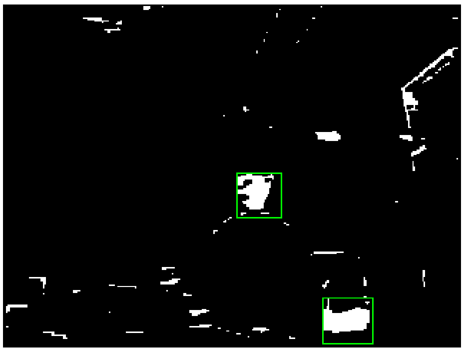


FIGURE 3- Skin-color regions.

FIGURE 4- Detected face.

Face Direction Computation

The system detects the eyes (pupils) and the nostrils in the zoomed-in image. We use the feature extraction module in the face recognition software library by Toshiba [Fukui and Yamaguchi (4)] for this process. Then, the system measures the horizontal distance between the left pupil and the left nostril *dl* and that for the right side *dr* as shown in Figure 5. From these two values it determines the direction of the gaze (face). In actuality, the robot does not need to compute the accurate direction. It only needs to determine whether or not the person is looking at the robot. Since the camera has turned in the human's direction, the frontal face must be observed if the human is looking at the robot face. If the ratio between *dl* and *dr* is close to 1, the human face can be considered to be facing toward the robot.

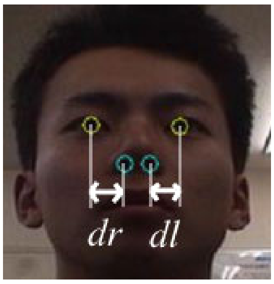


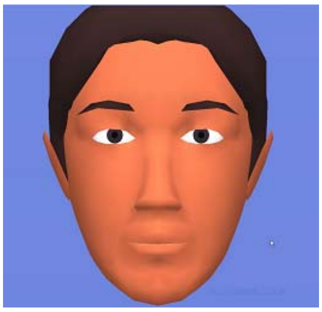
FIGURE 5- Face direction computation.

FACIAL ACTION OF THE ROBOT

We use the embodied agent developed by Hasegawa et al. [Hasegawa et al (6)] as our robot's head and face.

The CG head on the display turns in the same way as the camera as shown. Thus, the head keeps moving while the robot is trying to detect human faces. This solves the so-called Mona Lisa effect. In addition to the face direction, the distance between the two pupils of the robot face changes using the focused distance data of the camera as shown in Figure 6. The auto-focus module of the camera outputs the distance where the focus is adjusted. If the distance is large, the pupil distance is increased (Figure 6 (a)). If small, it is decreased (Figure 6 (b)). The face direction and the pupil distance show humans where the robot is looking. Such expressions by the robot are useful for humans to know when they can start communicating with the robot.

When the robot detects a human face, it turns its body in the human's direction. The CG head also turns back to the frontal position. The first condition necessary for eye contact is satisfied by this action. If the human still maintains his/her face directed toward the robot, the robot considers that the human is looking at it. The robot notices the human's gaze upon it. Since the movement of the robot's eyes stops, the human may in turn feel the robot's gaze on him/her. To make this feeling much clearer, the robot changes its facial expressions.



(a) When looking at near objects.

(b) When looking at far objects.

FIGURE 6- Changes of the pupil distance.

EYE CONTACT EXPERIMENTS

We performed experiments to examine whether or not our method would make humans feel that they had made eye contact with the robot.

First, we checked the effect of moving the head. We prepared two display cases: a moving head and a still image of the head with the frontal face. We used ten subjects; all were graduate students in our department. We asked them to give a value ranging from 0 (they do not perceive the gaze of the robot at all) to 6 (they do definitely) for each case.

Figure 7 shows the result. Comparing the ten resultant pairs with the Wilcoxon signed rank test gives a p-value of 0.0039. The two conditions are significantly different at the $p<0.05$ level. Therefore, moving the head was proven to be effective in making humans not perceive any eye contact from the robot.

Next, we performed an eye-contact experiment. We asked the same ten subjects to turn their head to look at the CG face and make eye contact with the robot, and then to give a subjective value from 0 (they do not perceive eye contact with the robot at all) to 6 (they definitely do) for each of the following three methods.

Method 1 (proposed method): The CG head is moving. When it notices that the human is looking at it, it stops in the frontal face position, then smiles.

Method 2: The CG head does not move and remains in the frontal face position. When it notices that the human is looking at it, it smiles.

Method 3: The CG head is moving. When it notices that the human is looking at it, it stops in the frontal face position without any changes in facial expressions.

Figure 8 shows the result. The Friedman test gives a p-value of 0.0016, showing that there are significant differences among the methods ($p<0.05$). The Scheffé test shows that Methods 1 and 3 give significantly different results ($p<0.05$). These results indicate that the proposed method (Method 1) is effective to make humans perceive eye contact with the robot.

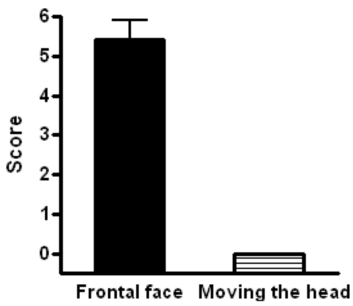


FIGURE 7- Effect of moving the head.

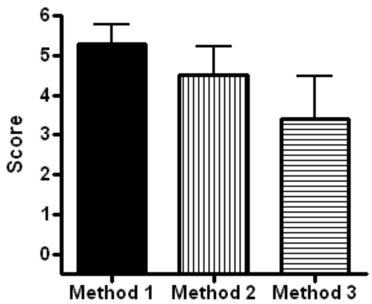


FIGURE 8- Eye-contact experiment.

Furthermore, we performed an experiment to examine what facial expression is most effective. We used Method 1 in the above experiment but tried changing the facial expression part. We examined the following three expressions: smiling (the same as in Method 1), nodding, and surprised (opening the eyes wide and raising the eyebrows). We used 11 subjects; all were graduate students in our department. We asked them to arrange these three expressions in order of making them perceive eye contact. We allowed them to give the same ranking to multiple expressions.

Figure 9 shows the result. (Note that the value 1 means the most effective expression for eye contact.) The Friedman test gives a p-value of 0.0057, showing that there are significant differences among these three ($p<0.05$). The Scheffé test shows that significant differences exist between the smiling and surprised expressions, and between the nodding and surprised expressions ($p<0.05$). We cannot say what is the best expression since there are various possible facial expressions. From these experiments, however, it can be seen that simple natural expressions such as smiling and nodding are good enough to tell humans that the robot is aware of their gaze.

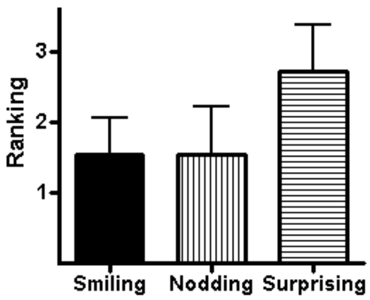


FIGURE 9- Expression experiment.

EYE CONTACT ROBOT WITH A WIDE FIELD OF VIEW

Experimental results have confirmed the effectiveness of the proposed eye-contact method. However, we noticed a problem with our first robot. It sometimes needs much time to find a person looking at the robot. The visual sensor of the robot is a pan-tilt-zoom controllable camera. Its field of view is narrow. The robot needs to rotate the camera to find a face candidate, and once it detects a candidate, it also needs to change the zoom. These mechanical actions require time. The human who would like to start communication with the robot should wait without moving his/her head until the robot completes all facial image processing. If he/she moves during the process, the robot may judge wrongly and start face detection again from scratch.

We have modified our robot to solve this problem. We have added two more cameras as shown in Figure 10. We have allocated a PC for image processing of each camera. If the robot detects a face candidate from one of the three cameras, the robot turns in the direction and uses the center camera for later processing.

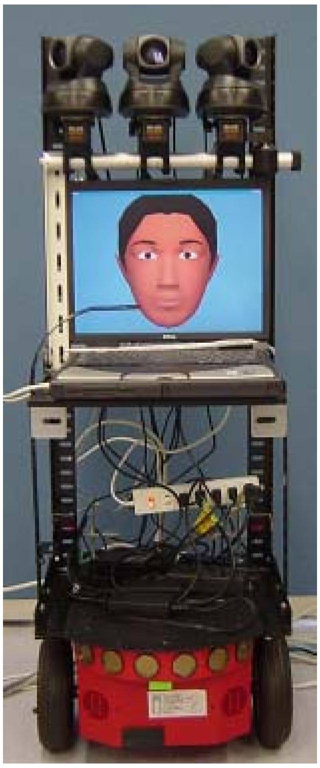


FIGURE 10- Robot with three cameras.

ROBOT THAT CAN BE CALLED BY EYE CONTACT

We have programmed the robot so that it will approach the person if he/she makes eye contact with it. In other words, we can call the robot just by making eye contact with it. The experiments described before show that the robot can do so when a person exists who really would like to make eye contact with the robot. A problem is, however, that the robot may start moving when people do not have such intention. We performed experiments to examine the capability of our eye contact method from this point.

The responsiveness of the robot to human's gaze can be controlled by the minimum gaze fixing time of the human that the robot needs to judge that he/she would like to make eye contact with the robot. The time is divided into two parts t_A and t_B. The former is the minimum time for the robot to determine that the human is looking in the direction of the robot. If the robot finds a face toward the robot and its direction holds for time $t_{A,}$ the robot turns toward the direction of the face. Then, if the face keeps looking at the robot for longer than time $t_{B,}$ the robot judges that the human would like to make eye contact.

We performed an experiment to determine appropriate t_A and $t_{B.}$ We asked a subject to sit in such a way that he/she faced away the robot by small angle $\theta = 5$ (degrees) as shown in Figure 11. He/she kept sitting for five minutes while pretending to eat and drink. (We simulated the robot waiter case described in the Introduction.) We counted the number of times that the robot erroneously judged that the subject would like to make eye contact. We did this experiment for four subjects; all were students in our department. We examined three parameter sets for t_A and $t_{B:}$ 0.5 and 0.25 seconds; 1.0 and 0.5; and 1.5 and 0.75.

Figure 12 shows the total number of errors. The robot made many errors. However, the error does not reduce much even though we increase the parameter values from 1.0 and 0.5 to 1.5 and 0.75. Thus, we have determined to use the parameters 1.0 and 0.5 for the robot in later experiments.

Next, we performed experiments to examine the effect of angle θ. We used the same experimental set-up as in the previous experiment. We used three subjects; they were students in our department. We asked them to pretend to eat and drink for five minutes and to try to make eye contact with the robot about in every minute during the period. This means that each subject tried to make eye contact five times in each session. We did the above for five angles θ, 0, 5, 10, 15, and 20 degrees. We counted the number of times that the robot did not approach the subjects even though they tried to make eye contact, and the number of times that the robot erroneously thought to be called by the subjects thorough eye contact.

Figure 13 shows the results. The figure indicates the total numbers of errors for the three subjects. The robot responded correctly all the time when the subjects tried eye contact. The false alarm type errors that the robot moved against human intentions reduce as angle θ increases. There were no errors when θ was 20 degrees. This means that if the robot is waiting while facing away from the humans by more than 20 degrees, it can respond correctly to their commands through eye contact.

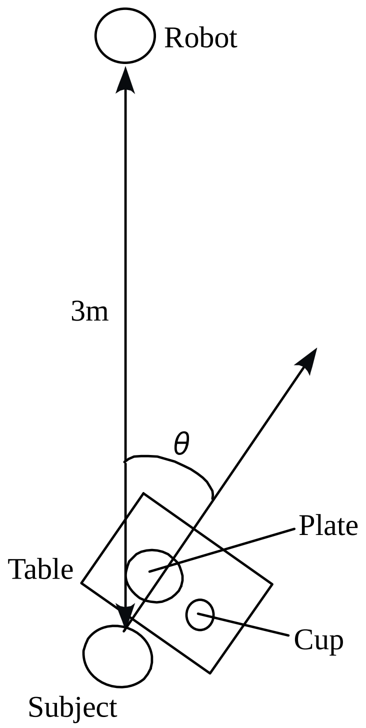


FIGURE 11- Positional relationships between the subject and the robot.

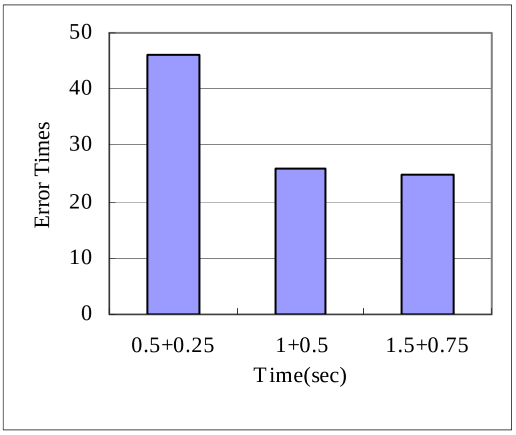


FIGURE 12- Relationships between the number of errors and the time thresholds.

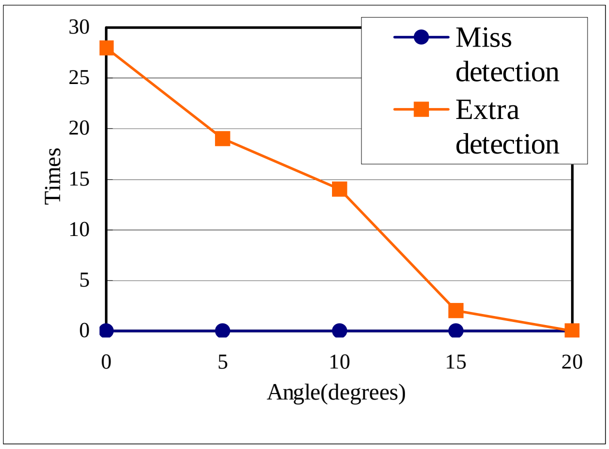


FIGURE 13- Number of errors and the facing angle between the human and the robot.

CONCLUSION

We have proposed an eye contact method to facilitate better communication between humans and robots. In addition, we have developed a robot system that we can beckon just by making eye contact. Experimental results suggest that we can develop robot waiters or clerks as described in the Introduction. The robots detect the humans in the room. They keep their positions so that their face directions can be more than 20 degrees away from those of all humans. In such situations, if anyone makes eye contact with a robot, the robot will come to the person. If there are many people in the room, it may be difficult for the robots to avoid facing all people at the same time. Occasional movements of the robots might solve the problem. We will investigate this problem and develop a robot system that can respond to natural human nonverbal behaviors.

ACKNOWLEDGMENTS

This work was supported in part by the Ministry of Education, Culture, Sports, Science and Technology under the Grant-in-Aid for Scientific Research (KAKENHI 14350127).

REFERENCES

１. Argyle, M., and Ingham, R., 1972, "Mutual gaze and proximity", Semiotica, 6, 32-49

２. Brooks, R.A., Breazeal, C., Marjanovic, M., Scassellati, B., and Williamson, M.M., 1988, "The cog project: Building a humanoid robot", Nehaniv, C. (ed.), Computation for Metaphors, Analogy and Agents, Lecture Notes in Artificial Intelligence, vol.1562, Springer-Verlag, 52-87

３. Cranach, M., 1971, "The role of orienting behavior in human interaction", Esser, A.H. (ed.), Behavior and Environment, Plenum Press, 217-237

４. Fukui, K. and Yamaguchi, O., 1998, "Facial feature point extraction method based on combination of shape extraction and pattern matching", Systems and Computers in Japan, 29, 6, 49-58

５. Hansen, J.P., Anderson, A.W., and Roed, P., 1995, "Eye-gaze control of multimedia systems", Anzai,Y., Ogawa, K., and Mori, H. (eds.), Symbiosis of Human and Artifact, vol.20A, Elsevier Science, 37-42

６. Hasegawa, O., Sakaue, K., Itou, K., Kurita, T., Hayamizu, S., Tanaka, K., and Otsu, N., 1997, "Agent oriented multimodal image learning system", Proc. IJCAI-WS IMS 1997, 29-34

７. Jacob, R.J.K., 1991, "The use of eye movements in human-computer interaction techniques: What you look at is what you get", ACM Trans. Information Systems 9, 3, 152-169

８. Kanda, T., Ishiguro, H., Ono, T., Imai, M., and Nakatsu, R., 2002, "Development and evaluation of an interactive humanoid robot “Robovie”", Proc. IEEE ICRA 2002, 1848-1855

９. Kendon, A., 1967, "Some functions of gaze direction in social interaction", Acta Psychologica, 26, 22-63

１０. Matsusaka, Y., Kubota, S., Tojo, T., Furukawa, K., and Kobayashi, T., 1999, "Multi-person conversation robot using multi-modal interface", Proc. SCI/ISAS, vol.7, 450-455

１１. Miyauchi, D., Sakurai, A., Nakamura, A. and Kuno, Y., 2004, "Active eye contact for human-robot communication", CHI2004 Extended Abstracts, 1099-1102

１２. Ohno, T. and Mukawa, N., 2003, "Gaze-based interaction for anyone, anytime", Proc. HCI International 2003, vol.4, 1452-1456