Active Eye Contact for Human-Robot Communication

Dai Miyauchi **Arihiro Sakurai** **Akio Nakamura** **Yoshinori Kuno**
Department of Information and Computer Sciences, Saitama University
255 Shimo-Okubo, Sakura-ku, Saitama-shi, Saitama 338-8570 JAPAN
{miyauchi, sakurai, nakamura, kuno}@cv.ics.saitama-u.ac.jp

Abstract
Eye contact is an effective means of controlling communication for humans, such as starting communication. It seems that we can make eye contact if we look at each other. However, this alone cannot complete eye contact. In addition, we need to be aware of being looked by each other. We propose a method of active eye contact for human-robot communication considering both conditions. The robot changes its facial expressions according to the observation results of the human to make eye contact. Then, we present a robot that can recognize hand gestures after making eye contact with the human to show the effectiveness of eye contact as a means of controlling communication.

Categories & Subject Descriptors: H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *Interaction styles*

General Terms: Design; Experimentation; Human Factors

Keywords: Eye contact; gaze; human-robot interface; nonverbal behavior; gesture recognition; embodied agent

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 humans and machines. Most of them use eye movements for pointing instead of a mouse or a joystick such as choosing an icon [5,7,11]. However, unlike arms and hands, eyes are not appropriate for precise pointing. We move our eyes in the direction where the objects exist that we would like to see. 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 using gaze for meta-communication. ROBITA [9] turns to the person speaking in conversation by multiple people. Robovie [8] and Cog [2] are other such examples.

Eye contact is a phenomenon that happens when two people cross their gaze. Since we perceive eye contact clearly, eye contact has stronger meta-communication capability than 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 in our direction. When he turns, our eyes meet his eyes. We make eye contact. Then, we move our hand slightly. This small gesture is enough to ask him to come.

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

Then, we present a robot that accepts simple hand-gesture commands after making eye contact with the human like a waiter in a restaurant as mentioned before. Hand gestures are good means of nonverbal communication. However, it is difficult to reduce false detection if we use simple hand movements for gesture commands, because such simple movements are frequently observed in other human activities than making the orders. Combining eye contact and gesture can solve the problem. We demonstrate this through experiments.

ROBOT SYSTEM
This section briefly shows our robot system used for eye-contact experiments before describing our method of making eye contact between humans and robots. Figure 1 shows the robot. We use a 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 may not attract humans' attention and they can concentrate on looking at the face on the computer display.

EYE CONTACT BETWEEN HUMANS AND ROBOTS
As mentioned in Introduction, two conditions: the gaze-crossing and the 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.

Figure 1. Eye-contact robot.

The second condition is difficult to be satisfied. Even if the robot has noticed that the human is looking at it, he/she may not be aware of this fact. We solve this problem by making the robot show explicitly this fact by changing its facial expressions. If the robot finds a human looking at it, it turns its face toward him/her. If he/she is still looking at it after this action, the robot assumes that the human is really looking at it. Then, it changes the facial expressions, such as smiling, to let him/her 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 not. There is, however, so-called the Mona Lisa effect. We feel that a face figure 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.

We have developed a nonverbal communication robot using the experimental system described in the previous section to examine the usefulness of the eye-contact method described above. The robot accepts hand gestures only after it made eye contact with the person making the gestures. The robot works as follows.

1. While rotating the camera, the robot detects face candidate regions. The CG head on the display turns in the same direction as the camera.
2. If it detects a face candidate, it 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, it turns in the direction of the human. It also turns the CG head to the frontal face position. If it observes that the face direction of the human is still toward it after this action, the robot changes the facial expressions of the CG head. (If the face direction is not toward the robot, return to 1.)
4. If the human moves a hand then, the robot considers the action as a meaningful gesture. If it can recognize the gesture, it makes the CG head nod to tell the human that it has understood the gesture.

FACIAL IMAGE PROCESSING

Psychological studies show that humans may avoid eye contact when they are too closely located. The frequency of eye contact increases as the distance between the humans increases [1]. This means that the robot should be able to make eye contact 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 is zoomed 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. First, skin color regions are extracted. 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. 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 in indoor scenes at a distance of 6 m.

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 [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 features *dr* as shown in Figure 2. It determines the gaze (face) direction from these two values. Actually, it does not need to compute the accurate direction. It only needs to determine whether or not the human 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 face direction can be considered to be toward the robot.

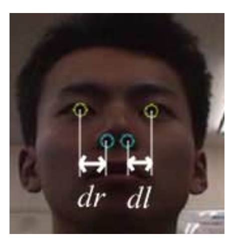

Figure 2. Face direction computation.

FACIAL ACTION OF THE ROBOT

We use the embodied agent developed by 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. Thus, the head keeps moving while the robot is trying to detect human faces. This solves so-called the 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. 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. If small, it is decreased. 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 the body in the human direction. The CG head also turns back into the frontal position. The first condition necessary for eye contact is satisfied by this action. If the human still holds his/her face direction toward the robot, the robot considers that the human is looking at the robot. The robot notices the human's gaze on it. Since the movement of the robot's eyes stops, the human may feel the robot's gaze on him/her. To make this feeling much clearer, the robot changes its facial expressions.

GESTURE RECOGNITION

We have adopted hand gestures as a communication means to examine the usefulness of eye contact. It seems to be easy to recognize simple hand gestures, such as the one that we use to express the meaning of "Come here." However, it is not so easy in actual complex scenes because such simple hand movements are often observed even though humans do not intend to do the gesture. Eye contact can be effective means to set up the communication channel between the human and the robot in such cases. The robot accepts the hand movements as a meaningful gesture only after making eye contact with the human. This can reduce false detection cases (false alarm errors).

We use the spotting recognition method based on continuous dynamic programming for gesture recognition [10]. We track a moving skin-color region and use the motion vector as the feature for recognition.

EXPERIMENTS

Eye Contact Experiments

We performed experiments to examine whether or not our method could make humans feel that they 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 3 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 is proved to be effective for the robot not to make humans perceive eye contact.

Then, 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 do definitely) 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 at the frontal position, then smiling.

Method 2: The CG head does not move with the frontal face. 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 at the frontal position without any facial expression changes.

Figure 4 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.

Furthermore, we performed an experiment to examine what facial expression was most effective. We used Method 1 in the above experiment by changing the facial expression part. We examined the following three expressions: smiling (the same as in Method 1), nodding, and surprising (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

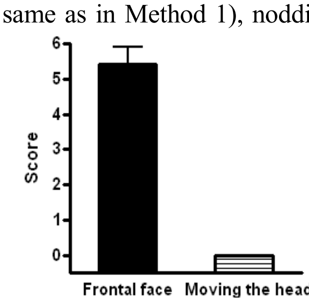

Figure 3. Effect of moving the head.

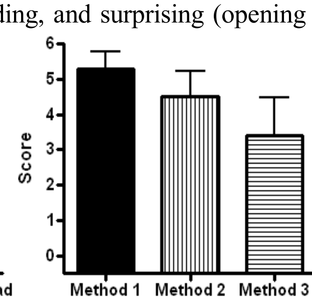

Figure 4. Eye-contact experiment.

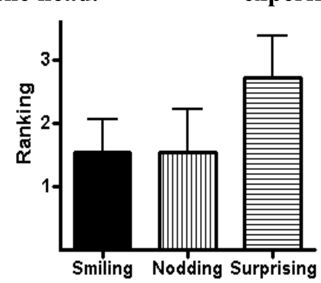

Figure 5. Expression experiment.

perceive eye contact. We allowed them to give the same place to multiple expressions.

Figure 5 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 surprising cases, and between the nodding and surprising cases ($p<0.05$). We cannot say what is the best expression since there are various possible facial expressions. From these experiments, however, simple natural expressions such as smiling and nodding are good enough to tell humans that the robot is aware of their gaze.

Gesture Recognition Experiments

We registered the hand gesture that we usually use to ask a person to come. First, we performed experiments to examine whether or not the robot was able to recognize the gesture. We did 20 experiments. It was able to recognize the gesture after making eye contact with the human in all cases. Then, we performed experiments to examine the effectiveness of eye contact in terms of avoiding false detection errors. We asked a subject to repeat the following action five times: walking from left to right (or vice versa) in the robot's camera field of view and returning to the start position. We consider this as a session and asked the subject to do 20 sessions in each of the following three cases.

(a) With eye contact: The robot starts gesture recognition only after making eye contact with the human.

(b) With face detection: It starts gesture recognition after detecting the face.

(c) None: It is ready for gesture recognition all the time.

Table 1 shows the experimental results. As shown in the table, the combination of eye contact and gesture proves to be effective to avoid false detection. The gesture recognition method used here is based on the movements of a skin-color region. Since the hand movement in the coming-here gesture is simple, the robot erroneously considered the face movement when the subject turned left or right as the gesture.

Table 1. Number of false detection times

	With eye contact	With face detection	None
False detection times	0	5	7

CONCLUSION

In this paper, we have proposed an active eye contact method between humans and robots. Then, we have presented a robot that can recognize hand gestures after making eye contact with the human. Experimental results show the effectiveness of eye contact as a means of meta-communication.

Eye contact can be useful in various other occasions than the gesture recognition case. We are now working on this issue. The current method assumes a person at a time. This is also left for future work.

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, 15017211).

REFERENCES

1. Argyle, M., and Ingham, R., Mutual gaze and proximity. *Semiotica 6* (1972), 32-49.
2. Brooks, R.A., Breazeal, C., Marjanovic, M., Scassellati, B., and Williamson, M.M., 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 (1998), 52-87.
3. Cranach, M., The role of orienting behavior in human interaction. Esser, A.H. (ed.), *Behavior and Environment,* Plenum Press (1971), 217-237.
4. Fukui, K. and Yamaguchi, O., Facial feature point extraction method based on combination of shape extraction and pattern matching. *Systems and Computers in Japan 29*, 6 (1998), 49-58.
5. Hansen, J.P., Anderson, A.W., and Roed, P., Eye-gaze control of multimedia systems. Anzai,Y., Ogawa, K., and Mori, H. (eds.), *Symbiosis of Human and Artifact*, *vol.*20A, Elsevier Science (1995), 37-42.
6. Hasegawa, O., Sakaue, K., Itou, K., Kurita, T., Hayamizu, S., Tanaka, K., and Otsu, N., Agent oriented multimodal image learning system. *Proc. IJCAI-WS IMS 1997*, 29-34.
7. Jacob, R.J.K., The use of eye movements in human-computer interaction techniques: What you look at is what you get. *ACM Trans. Information Systems 9*, 3 (1991), 152-169.
8. Kanda, T., Ishiguro, H., Ono, T., Imai, M., and Nakatsu, R., Development and evaluation of an interactive humanoid robot "Robovie". *Proc. IEEE ICRA 2002*, 1848-1855.
9. Matsusaka, Y., Kubota, S., Tojo, T., Furukawa, K., and Kobayashi, T., Multi-person conversation robot using multi-modal interface. *Proc. SCI/ISAS 1999*, *vol.*7, 450-455.
10. Nishimura, T., Mukai, T., and Oka, R., Spotting recognition of gestures performed by people from a single time-varying image. *Proc. IEEE/RSJ IROS 1997*, *vol.*2, 967-972.
11. Ohno, T. and Mukawa, N., Gaze-based interaction for anyone, anytime. *Proc. HCI International 2003*, *vol.*4, 1452-1456.