

Image-Based Active Control for AEM Function of ARM-COMS

Teruaki Ito^{1(✉)} and Tomio Watanabe²

¹ Tokushima University, 2-1 Minami-Josanjima, Tokushima 770-8506, Japan
tito@tokushima-u.ac.jp

² Okayama Prefectural University, 111 Tsuboki, Souja, Okayama 719-1197, Japan
watanabe@cse.oka-pu.ac.jp

Abstract. This study has proposed an idea of remote individuals' connection through augmented tele-presence systems called ARM-COMS (ARm-supported eMbodied COMmunication Monitor System). ARM-COMS is composed of a tablet PC and a desktop robotic arm. The table PC in ARM-COMS is a typical ICT (Information and Communication Technology) device and the desktop robotic arm works as a manipulator of the tablet. ARM-COMS has three types of functions; namely, autonomous positioning (AP), autonomous entrainment movement (AEM), and autonomous entrainment positioning (AEP), which are the key to connect the remote individuals. This paper mainly focuses on AEM function, which was implemented using a three-steps of control approach, including face detection, landmark detection and face orientation estimation. Reviewing the experimental results using a prototype system based on the three-step approach, the feasibility of this control procedure will be discussed.

Keywords: Embodied communication · Augmented tele-presence robotic arm manipulation · Face detection · Landmark detection · Face orientation

1 Introduction

ICT (Information and Communication Technology) technologies has the potential to further enhance a good communication. TV phone is regarded as a tool in SF movies in the old days. However, the time changed so fast that WiFi based video communication tool is one of the convenient tools which can be available to any of us [1]. On the other hand, it addresses the two types of critical issues, or the lack of tele-presence feeling and the lack of relationship feeling in communication [4].

One of the solutions to the former issue is proposed by several ideas of robot-based remote communication systems; such as physical telepresence robots [8, 21], or an idea of anthropomorphization [13]. Distance communication is supported by the basic functions of physical tele-presence robots, such as face image display of the operator [14], as well as tele-operation function such as remote-drivability to move around [9], or tele-manipulation [9]. However, it is recognized that a gap between robot-based video communication and face-to-face one have not been narrowed yet.

Recently, a new challenge has been undertaken by an idea of robotic arm-typed systems [24]. For example, Kubi [11], which is a non-mobile arm type robot, allows the remote user to “look around” during their video communication by way of commanding Kubi where to aim the tablet with an intuitive remote control over the net. Furthermore, an idea of enhanced motion display has also been reported [15] to show its feasibility over the conventional display. However, the movement of human body as a non-verbal message from a remote person is still an open issue on robotic arm-typed systems.

Considering the physical entrainment motion in human communication [23], this research challenges the two issues, which are the lack of tele-presence feeling and the lack of relationship in communication [5]. This paper focuses on AEM function, which was implemented using a three steps of control approaches, including face detection, landmark detection and face orientation estimation, of which procedure is also presented in this paper. Reviewing the experimental results using a prototype system, the feasibility of this control procedure will be discussed.

2 Overview of ARM-COMS(ARM-Supported eM-bodied Communication Monitor System)

2.1 System Overview

ARM-COMS is composed of a tablet PC and a desktop robotic arm. The table PC in ARM-COMS is a typical ICT (Information and Communication Technology) device and the desktop robotic arm works as a manipulator of the tablet, of which position and movements are autonomously manipulated based on the behavior of a human user. This autonomous manipulation is controlled by the head movement of a master person, which can be recognized by a portable device, such as a Kinect [10] sensor, or a general USB camera, and its detected signals are transferred to the PC under which ARM-COMS is controlled.

A prototype system of ARM-COMS is shown in Fig. 1, where a 5-DOF robotic arm is the key manipulator which controls the movement of table PC. The robotic arm can be controlled by the acceleration sensor attached to the human subject [6]. Feasibility experiments in remote communication with/without ARM-COMS were conducted and its positive effects were recognized. Since the physical sensor was attached to the human subject in this prototype, non-contact type of motion sensor was also tested based on the hand gesture manipulation [7]. Using a hand gesture, which is based on more wide range of movement, motion control of ARM-COMS was analyzed. As the result of experiments, ARM-COMS could mimic the hand gesture motion based on the motion sensor control, using the three types of hand gestures which shows nodding, head-shaking, and head-tilting of head gesture.

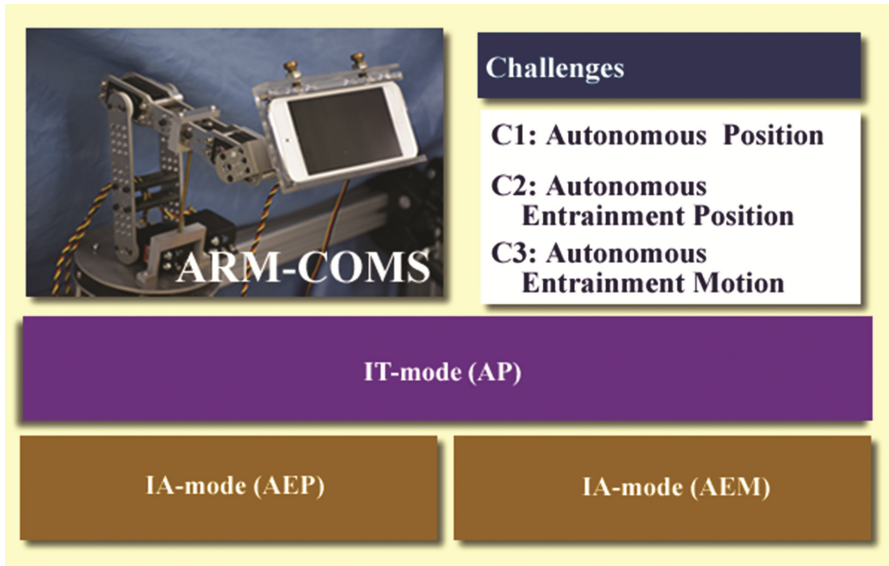


Fig. 1. Overview of ARM-COMS

2.2 IT/IA Modes with AP/AEM/AEP Functions of ARM-COMS

ARM-COMS operates as an intelligent ICT system in *IT-mode* as well as an intelligent avatar system in *IA-mode* (Fig. 1).

Considering the user's physical position, AP (Autonomous Position) function of *IT-mode* enables the tablet PC autonomously and automatically to take an appropriate position towards the user (Challenge 1). For example, ARM-COMS moves the tablet PC closer to the user when a phone call comes in.

ICT devices allow us to communicate with others over the network. However, there is a significant difference in communication between the face-to-face communication and the video communication. It is reported that sharing the same physical space and atmosphere among the participants typically plays a key role in human communication. Since the entrainment is associated with physical movement of a person [20], AEM (Autonomous Entrainment Motion) function of *IA-mode* enables physical movement of the tablet PC during remote communication for entrainment acceleration by mimicking the head movement of its master person (Challenge 2).

In addition to the head movement of a remote person, AEP (Autonomous Entrainment Position) function of *IA-mode* enables to express of relationship between the persons by the determination of appropriate distance (Challenge 3).

3 Image-Based Motion Control of ARM-COMS

A prototype of ARM-COMS system has been developed to study the feasibility of the head motion control based on the video image [18]. The prototype system adopts a table-top 5 DOF robotic arm, which is controlled by a microcontroller using simple commands based on gesture signals from an USB camera. The prototype is designed to mimic the basic human head motion as the AEM function, which is one of the challenges of ARM-COMS as mentioned in the previous session.

Previous paper [7] reports the hand gesture motion using ARM-COMS based on the finger motion sensor control [12], where ARM-COMS mimics the hand gesture based on the signals from the sensor. This type of finger motion control sensor is a non-contact type and easy to use. However, it is not so convenient for the user to prepare a special sensor, which is required to detect the hand motion. Therefore, in this experiments, head motion of a human subject was traced by a general USB camera, which generates the control code of AMR-COMS to mimic the head motion. Focusing on the two types of typical human head gestures, namely, nodding motion for affirmative meaning and head shaking motion for negative meaning, the experimental setup was designed as shown in Fig. 2.

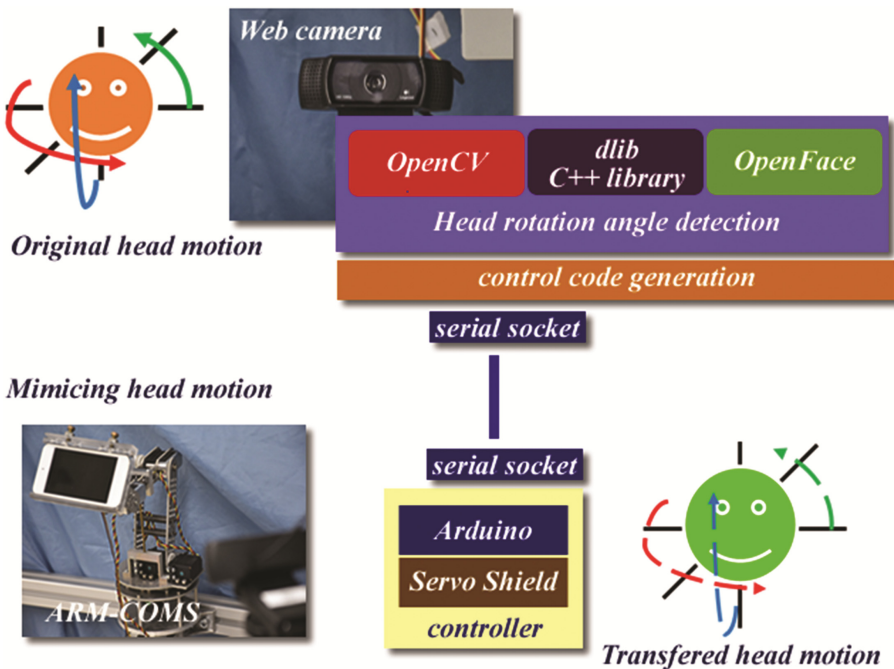


Fig. 2. Configuration of ARM-COMS prototype

Face detection procedure of a prototype of ARM-COMS is based on the algorithm of FaceNet [19], which includes image processing library OpenCV 3.1.0 [16], machine learning library dlib 18.18 [2], and face detection tool OpenFace [17] which were installed on a control PC with Ubuntu 14.04 [22] as shown in Fig. 2. Using the input image data from USB camera, landmark detection is processed.

The control procedure is based on the following three steps, which include the face detection, the landmark detection and the face orientation estimation.

The face detection narrows the face area captured by the USB camera. Using an extracted face image by open CV library [16], outline contour of the face is generated as shown in Fig. 3, which is used in landmark detection in the second step.

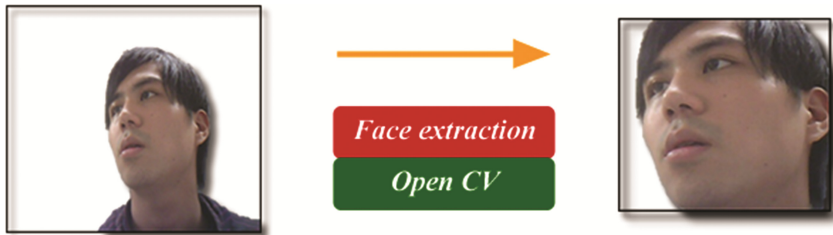


Fig. 3. Face extraction

The landmark detection captures the 68 landmarks defined the face as shown in Fig. 4. This facial landmark detection of 68 points is detected using dlib library [2].

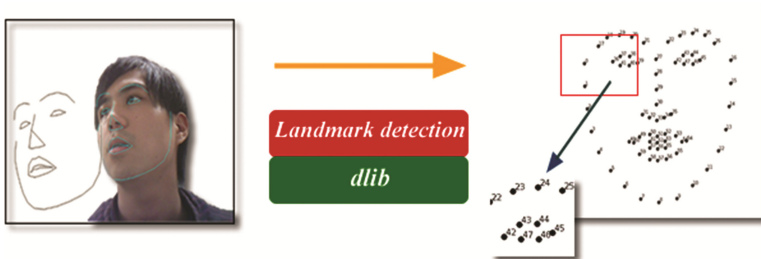


Fig. 4. Landmark detection

The face orientation estimation determines the orientation of the face. OpenFace [17] enables the calculation of rotational angles of roll-pitch-yaw by three dimensional coordinates of landmark points based on a perspective n-point (PnP) techniques as shown in Fig. 5.

Then control codes using roll-pitch-yaw angles are sequentially generated and transferred to ARM-COMS via serial sockets connecting the PC and the controller.

Head-motion of a human subject is detected and traced according to the movement of the subject. In order to evaluate the traceability of ARM-COMS, motion control



Fig. 5. Face orientation

experiments were conducted. Figure 6 shows the snapshot of experimental set-ups. A magnetic receiver A (Fastrak RX-2 [3]) is attached to the head of human subject and another magnetic receiver B is attached to the ARM-COMS. The movements of the head motion and ARM-COMS motion were detected simultaneously and recorded through the magnetic transmitter (Fastrak TX-2 [3]). A USB camera (Buffalo BSW20K04H) captures the image of human subjects during the experiments. A desktop PC (Windows 7/64) was used for the data collection, whereas a laptop pc (Ubuntu 14.04) was used for ARM-COMS control.

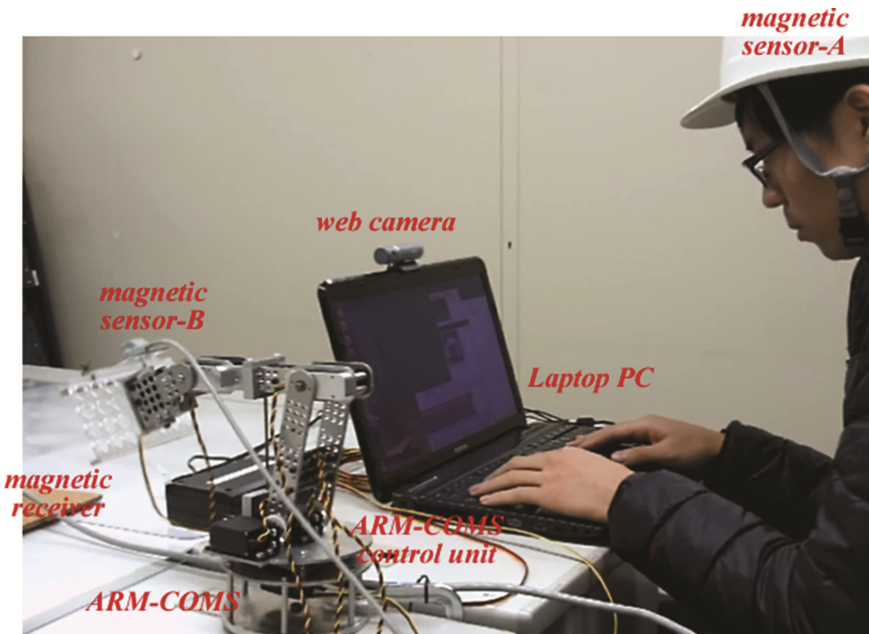


Fig. 6. Experimental setup

Experimental procedure:

- Step 1: Three nodding with over a period of one second each, followed by another three nodding with over a period of two seconds each, and another three nodding with over a period of three seconds each.
- Step 2: Three head shaking with over a period of one second each, followed by another three head shaking with a period of two seconds each and another three head shaking with a period of three seconds each.

Figure 7 shows the nodding gesture in head motion. The red line shows the time series variation of head angles in nodding gesture conducted three times in a consecutive manner, whereas the green line shows time series variation of pitch angles of the corresponding ARM-COMS motion. The graph shows that ARM-COMS mimics the head nodding motion quite well.

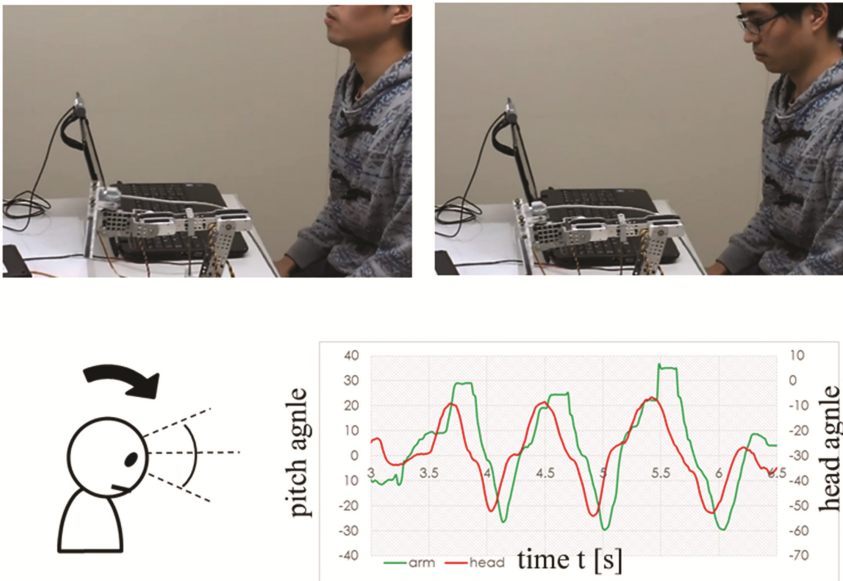


Fig. 7. Analysis of nodding motion (Color figure online)

Figure 8 shows the head shaking gesture of head movement. The red line shows the time series variation of head shaking gesture conducted three times in a consecutive manner, whereas the green line shows the time series variation of corresponding ARM-COMS motion. As seen from the graph, the ARM-COMS mimics the head shaking motion very smoothly.

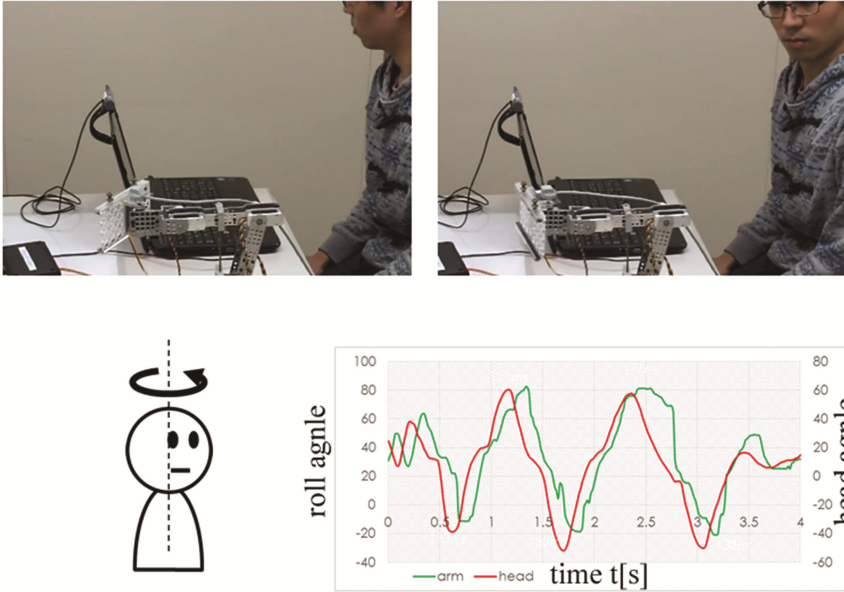


Fig. 8. Analysis of head shaking motion (Color figure online)

4 Concluding Remarks

This study proposes an idea of human-computer interaction through remote individuals' connection with augmented tele-presence systems called ARM-COMS (ARm-supported eM-bodied COMMunication Monitor System). This paper shows a prototype of active display monitor for ARM-COMS with the two types of modes, or IT-mode and IA-mode. Considering the three challenges using the basic functions, or AP, AEM and AEP functions, this paper focuses on image-based motion control of tablet PC held by ARM-COMS to evaluate the feasibility of AEM function. The results of experiment show the feasibility of head motion control based on the image-based active control for AEM function of ARM-COMS.

Acknowledgement. This work was supported by JSPS KAKENHI Grant Numbers JP16K00274, JP26280077. The author would like to acknowledge Hiroki KIMACHI, and all members of Collaborative Engineering Labs at Tokushima University, and Center for Technical Support of Tokushima University, for their cooperation to conduct the experiments.

References

1. Abowdm, D.G., Mynatt, D.E.: Charting past, present, and future research in ubiquitous computing. *ACM Trans. Comput. Hum. Interact. (TOCHI)* 7(1), 29–58 (2000)
2. Dlib c++ library. <http://dlib.net/>
3. FASTRK. <http://polhemus.com/motion-tracking/all-trackers/fastrak>
4. Greenberg, S.: Peepholes: low cost awareness of one's community. In: *Conference Companion on Human Factors in Computing Systems: Common Ground*, Vancouver, British Columbia, Canada, pp. 206–207 (1996)
5. Ito, T., Watanabe, T.: Three key challenges in ARM-COMS for entrainment effect acceleration in remote communication. In: Yamamoto, S. (ed.) *HCI 2014. LNCS*, vol. 8521, pp. 177–186. Springer, Cham (2014). doi:[10.1007/978-3-319-07731-4_18](https://doi.org/10.1007/978-3-319-07731-4_18)
6. Ito, T., Watanabe, T.: ARM-COMS for entrainment effect enhancement in remote communication. In: *Proceedings of the ASME 2015 International Design Engineering Technical Conferences & Computers and Information Engineering Conference (IDETC/CIE2015)*, August, Boston, USA, DETC2015-47960 (2015)
7. Ito, T., Watanabe, T.: Motion control algorithm of ARM-COMS for entrainment enhancement. In: Yamamoto, S. (ed.) *HIMI 2016. LNCS*, vol. 9734, pp. 339–346. Springer, Cham (2016). doi:[10.1007/978-3-319-40349-6_32](https://doi.org/10.1007/978-3-319-40349-6_32)
8. Kashiwabara, T., Osawa, H., Shinozawa, K., Imai, M.: TEROOS: a wearable avatar to enhance joint activities. In: *Annual conference on Human Factors in Computing Systems*, pp. 2001–2004 (2012)
9. Kim, K., Bolton, J., Girouard, A., Cooperstock, J., Vertegaal, R.: TeleHuman: effects of 3D perspective on gaze and pose estimation with a life-size cylindrical telepresence pod. In: *Proceedings of CHI2012*, pp. 2531–2540 (2012)
10. Kinect. <https://dev.windows.com/en-us/kinect>
11. Kubi. <https://www.revolverobotics.com>
12. Leapmotion. <https://www.leapmotion.com/>
13. Osawa, T., Matsuda, Y., Ohmura, R., Imai, M.: Embodiment of an agent by anthropomorphization of a common object. *Web Intell. Agent Syst. Int. J.* 10, 345–358 (2012)
14. Otsuka, T., Araki, S., Ishizuka, K., Fujimoto, M., Heinrich, M., Yamato, J.: A realtime multimodal system for analyzing group meetings by combining face pose tracking and speaker diarization. In: *Proceedings of the 10th International Conference on Multimodal Interfaces (ICMI 2008)*, Chania, Crete, Greece, pp. 257–264 (2008)
15. Ohtsuka, S., Oka, K.K., Tsuruda, T., Seki, M.: Human-body swing affects visibility of scrolled characters with direction dependency. In: *Society for Information Display (SID) 2011 Symposium Digest of Technical Papers*, pp. 309–312 (2011)
16. OpenCV. <http://opencv.org/>
17. OpenFace API Documentation. <http://cmusatyalab.github.io/openface/>
18. Sato, T., Kanbara, M., Yokoya, N., Takemura, H.: Dense 3-D reconstruction of an outdoor scene by hundreds-baseline stereo using a hand-held video camera. *Int. J. Comput. Vis.* 47(1–3), 119–129 (2002)
19. Schoff, F., Kalenichenko, D., Philbin, J.: FaceNet: a unified embedding for face recognition and clustering. In: *IEEE Conference on CVPR 2015*, pp. 815–823
20. Sirkin, D., Ju, W.: Consistency in physical and on-screen action improves perceptions of telepresence robots. In: *HRI 2012 Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, pp. 57–64 (2012)

21. Tariq, A.M., Ito, T.: Master-slave robotic arm manipulation for communication robot. In: Proceedings of 2011 Annual Meeting on Japan Society of Mechanical Engineer, Vol. 11(1), p. S12013, September 2011
22. Ubuntu. <https://www.ubuntu.com/>
23. Watanabe, T.: Human-entrained Embodied Interaction and Communication Technology. In: Fukuda, S. (ed.) Emotional Engineering, pp. 161–177. Springer, New York (2011)
24. Wongphati, M., Matsuda, Y., Osawa, H., Imai, M.: Where do you want to use a robotic arm ? And what do you want from the robot ? In: International Symposium on Robot and Human Interactive Communication, pp. 322–327 (2012)