Tele-Embodied Agent (TEA) for Video Teleconferencing

Muhammad Sikandar Lal Khan[†], Shafiq ur Réhman[*†], Zhihan LU[†], Haibo Li[‡]
[†]i2Lab, Dept. of Applied Physics and Electronics (TFE), UmeåUniversity,Umeå, Sweden
[‡]Royal Institute of Technology (KTH), Stockholm, Sweden
{muhammad.sikandar.lal.khan,shafiq.urrehman,zhihan.lu}@umu.se

ABSTRACT

We propose a design of teleconference system which express nonverbal behavior (in our case head gesture) along with audio-video communication. Previous audio-video conferencing systems are abortive in presenting nonverbal behaviors which we, as human, usually use in face to face interaction. Recently, research in teleconferencing systems has expanded to include nonverbal cues of remote person in their distance communication. The accurate representation of non-verbal gestures for such systems is still challenging because they are dependent on hand-operated devices (like mouse or keyboard). Furthermore, they still lack in presenting accurate human gestures. We believe that incorporating embodied interaction in video teleconferencing, (i.e., using the physical world as a medium for interacting with digital technology) can result in nonverbal behavior representation. The experimental platform named Tele-Embodied Agent (TEA) is introduced which incorperates remote person's head gestures to study new paradigm of embodied interaction in video teleconferencing. Our preliminary test shows accuracy (with respect to pose angles) and efficiency (with respect to time) of our proposed design. TEA can be used in medical field, factories, offices, gaming industry, music industry and for training.

1. INTRODUCTION & MOTIVATION

During conversation human use multimodal information; including not only verbal information but also prosodic and visual information [7]. Similarly, in face-to-face communication we take assistance of facial movements, facial expressions, gestures and gaze to convey better meaning to our conversation. Whereas, in Computer Mediated Communication (CMC) like video conferencing systems, the standard interaction approaches rely on text, audio and video. These standard video conferencing systems are abortive in expressing nonverbal behaviors which we as human usually employ in face to face interaction [5]. But, recently researchers are trying to make teleconferencing more interactive and more realistic, so that people sitting in a conference or meeting can feel presence of a remote person. In recent research, Mebot [1]and ESP [4] have new dimension for video teleconferencing through telerobots to convey social expression like hand gesture, head movement and proxemics. Hence, wide range of common human communications, interactions, and activities can be better captured and expressed between distant locations using internet-controlled telerobots which can act as a physical agent for collaborators.

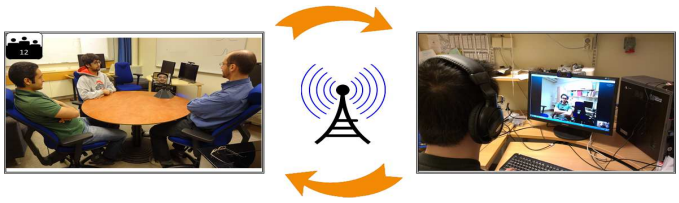

Figure 1: Application Scenario

We propose a new type of video conferencing system which differs mechanically and conceptually from present teleconferencing systems. The design of our prototype system is based on Embodied Interaction [3] concepts by which one can express remote person's head gestures with collaborators sitting in a meeting room. We named our telerobot as Tele-Embodied Agent (TEA); a telerobot which embodies the remote participant's head gestures and express these gestures among his/her collaborators at distant location. In addition to audio-video conferencing, TEA has the capability to mimic head pose (comprising all three degrees of head rotation) of remote participant to express iconic cues that people use in face to face interaction. Our telerobot is controlled by remote user's head motion instead of hand-operated devices. The system is proficient in expressing head gestures like turn taking, nodding and shaking. This raises involvement, and increases level of attention; hence making conversation more interesting. The application scenario is shown in Figure 1.

2. TELE EMBODIED AGENT (TEA)

We designed a prototype system named Tele-Embodied Agent (TEA). TEA system consists of two subsystems: Hardware Design Module & Software Design Module. For TEA, remote participant's head gestures are estimated and the telerobot moves accordingly among local collaborators. Tablet PC mounted on TEA is used for audio-video communication through Skype.

*corresponding author:shafiq.urrehman@umu.se

MUM '13, Dec 02-05 2013, Luleå, Sweden

Figure 2: (a) CAD Model (b) Mechanical Platform

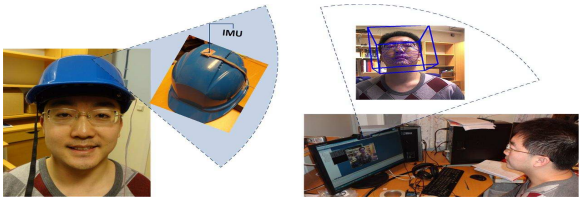

Figure 3: (a) IMU Based head pose estimation (b) CLM plus Geometric Based Sol.

2.1 Hardware Design Module

TEA hardware design is based on three degrees of freedom platform that actuates a tablet PC (or smart phone). For current design, we use three servo motors in configuration which mimic all three degrees of head movement (i.e. yaw, pitch and roll). The CAD model of our prototype is shown in figure 2 (a) and the original mechanical platform is shown in figure 2 (b).

2.2 Software Design Module

In software module, we used extended computer vision and machine learning algorithm using CLM [2] and geometric head pose estimation algorithm [6] to calculate head pose of remote person. Figure 3 (a) shows IMU based head pose values (ground truth values) and figure 3 (b) shows CLM plus geometric based solution for pose calculation. We validated pose values from our algorithm against head-mounted IMU values. We performed a laboratory test to estimate the strength of our TEA based on two parameters; namely - Accuracy with respect to pose angles and Efficiency with respect to time. The results show that average error is 0.183, 1.75, and 0.567 (degrees) in yaw, pitch and roll respectively as shown in figure 4. In figure 4, actual angles are from CLM plus geometric algorithm and Expected angles are from IMU. Finally, over all speed of our system is 10 Hz.

3. SYSTEM DEPLOYMENT

The system deployment scenario is as follow. The remote user sits in front of his/her PC equipped with camera while

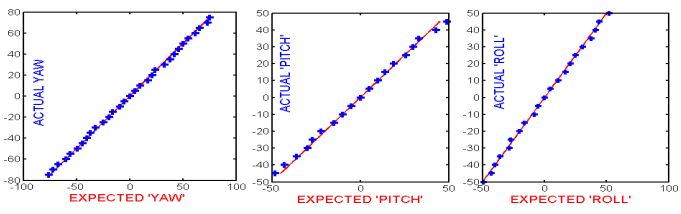

Figure 4: Expected Vs Actual angles for (a) Yaw (b) Pitch (c) Roll

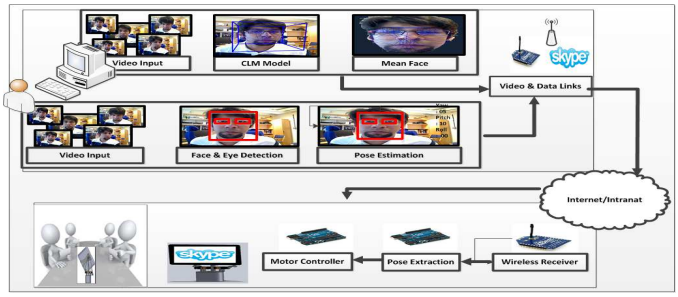

Figure 5: System Level Diagram

local collaborators are sitting in meeting/conference room. Audio and video communication between a remote person and his collaborators is established through a customized audio video broadcasting software. The data communication is established with XBEE wireless transceiver. Wireless commands from a remote user are sent to TEA (placed among local collaborators) and TEA micro-controller decodes these commands and actuates three degrees of freedom platform (TEA) according to Yaw, Pitch and Roll angles of the remote user. This is accomplished by giving PWM signal to each servo motor to move it to certain angle. The functional diagram of proposed system is shown in figure 5.

4. DISCUSSION & FUTURE WORK

We proposed a novel design for video teleconferencing system which mimics head gesture of remote participant. Tele-Embodied Agent (TEA) is inexpensive, portable and easy to use system. The integration of audio, video and mechanical embodiment of head gesture is novel approach. According to our best knowledge it is first time we considered all three rotational axes of our head (Yaw, Pitch and Roll). Our preliminary results shows accuracy and efficiency of proposed prototype system. Our future goals are to investigate the effect of precise head gesture on creating 'feeling of presence' and on improving 'quality of interaction'.

5. REFERENCES

[1] S. Adalgeirsson and C. Breazeal. Mebot: A robotic platform for socially embodied telepresence. In *ACM Press, HRI*, 2010.

[2] D. Cristinacce and T. Cootes. Feature detection and tracking with constrained local models. In *BMVC*, 2006.

[3] P. Dourish. *Where The Action Is: The Foundations Of Embodied Interaction*. The MIT Press, 2004.

[4] G. V. et. al. Embodied social proxy: Connecting hub-and-satellite teams. In *28th Int. Conf. Human Factors in Com. Sys.*, 2010.

[5] R. Kaliouby and P. Robinson. Real-time inference of complex mental states from facial expressions and head gestures. *Real-Time Vision for HCI*, 2005.

[6] M. Khan, Z. Lu, S. Réhman, and H. Li. Head orientation modeling: Geometric head posed estimation using monocular camera. In *1st IEEE/IIAE Int. Conf. Intelli. Sys. and Im. Pro.*, 2013.

[7] R. M. Krauss, Y. Chen, and P. Chawla. Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? In *M. Zanna (Ed.), Advances in Exp. Soc. Psy.*, 1995.