A System for Reconstructing Multiparty Conversation Field based on Augmented Head Motion by Dynamic Projection

Kazuhiro Otsuka
NTT CSL
3-1 Morinosato Wakamiya
Atsugi-shi, Japan
otsuka@ieee.org

Kamil Sebastian Mucha
NTT CSL
3-1 Morinosato Wakamiya
Atsugi-shi, Japan
kamil@cs.kecl.ntt.co.jp

Shiro Kumano
NTT CSL
3-1 Morinosato Wakamiya
Atsugi-shi, Japan
shiro.kumano@lab.ntt.co.jp

Dan Mikami
NTT CSL
3-1 Morinosato Wakamiya
Atsugi-shi, Japan
mikami.dan@lab.ntt.co.jp

Masafumi Matsuda
NTT CSL
3-1 Morinosato Wakamiya
Atsugi-shi, Japan
masafumi.matsuda@lab.ntt.co.jp

Junji Yamato
NTT CSL
3-1 Morinosato Wakamiya
Atsugi-shi, Japan
yamato@brl.ntt.co.jp

ABSTRACT
A novel system is presented for reconstructing, in the real world, multiparty face-to-face conversation scenes; it uses dynamics projection to augment human head motion. This system aims to display and playback pre-recorded conversations to the viewers as if the remote people were taking in front of them. This system consists of multiple projectors and transparent screens. Each screen separately displays the life-size face of one meeting participant, and are spatially arranged to recreate the actual scene. The main feature of this system is dynamics projection, screen pose is dynamically controlled to emulate the head motions of the participants, especially rotation around the vertical axis, that are typical of shifts in visual attention, i.e. turning gaze from one to another. This recreation of head motion by physical screen motion, in addition to image motion, aims to more clearly express the interactions involving visual attention among the participants. The minimal design, frameless-projector-screen, with augmented head motion is expected to create a feeling that the remote participants are actually present in the same room. This demo presents our initial system and discusses its potential impact on future visual communications.

Categories and Subject Descriptors
H1.2 [Models and Principles]: User/Machine System — Human Information Processing

General Terms
ALGORITHMS, DESIGN, HUMAN FACTORS

Keywords
face-to-face conversation, multimodal interaction, projector system, telepresence, visual attention, visual communication

1. INTRODUCTION
Face-to-face conversation is one of the most basic forms of communication in daily life and group meetings are used for conveying/sharing information, understanding others’ intention/emotion, and making decisions. As a part of visual communication research, this study focuses on the essential problem of how people perceive and understand others’ conversations, and how to design/built a system that allows outside people to re-experience the conversation by recreating the real face-to-face setting as closely as possible. We formulate the problem of conversation field reconstruction as the re-creation, in the real world, of multiparty face-to-face conversation scenes using novel devices with augmented expression modality.

2. DESIGN CONCEPT
Considering the importance of the nonverbal information exchanged in face-to-face conversations, such as facial expressions, eye-gaze, and head/body gestures, our system aims to reproduce it in an actual environment that closely mirrors the original one. Among the different forms of non-verbal information, this study focuses on the visual attention of conversation participants. Visual attention, also called gaze, can indicate “who is looking at whom”; it is particularly important in understanding the structure of a conversation, e.g. “who is talking to whom”. To express visual attention to the viewers, we newly introduce a dynamics projection as an augmented modality that indicates the direction and shift of visual attention of the participants. Details of the design concept are as follows.

First, this system consists of multiple screens, where each screen corresponds to a different individual; they are spatially configured to recreate the original spatial arrangement of the participants. This arrangement is one requirement for intuitively understanding the visual attention shift during the conversation. Fig. 1 shows an overview of the proposed system.

Second, each screen displays the (near) frontal view of the life-sized face and upper body of one participant. Unlike avatar-based systems, we preserve the original image of the face itself, which gives viewers the full range of nonverbal messages including subtle facial expressions.

Copyright is held by the author/owner(s).

ACM 978-1-4503-0616-4/11/11.
Third, our system employs flat transparent screens with back-projectors. Participant images are fused into the room background from the viewer's eye. The screen has no frames, edges, nor margin, unlike other display devices used for telecommunications. The frameless transparent screens create the impression that the people's faces are floating in the air.

Fourth, as the most notable feature, the screen dynamically changes its pose to emulate the head motion of each person. We place particular emphasis on head rotation around the vertical axis, i.e. head rotation when turning to one to another participant. This additional expression modality, augmented by physical screen motion, combines with original image details, aims to more clearly represent the changes in visual attention among the participants. This idea of head motion augmentation is based on the nature of human perception called biological motion and the theory of mind. These theories indicate that humans can anthropomorphize an object when it moves like a human. Also, the viewer's attention can be induced by screen motion in their peripheral vision, as in a real setting.

To summarize the above, our system features a minimal design approach, the simple and abstract display devices that are animated with realistic human motion. This tries to re-create the conversation scene 'as is', and allows the viewer to intuitively experience and understand the conversation.

3. SYSTEM CONFIGURATION

Fig. 1 overviews the proposed system. Fig. 1(a) shows actual conversation scenes, and Fig. 1(b) shows the reconstructed conversation scenes. The sensing part (left in Fig. 1(a)) includes multiple cameras and microphones, which capture the face images and voice of each participant. One example of the sensing part can be found in [1]. The visualization part consists of multiple projectors and screens with actuators to control screen motion.

The projector screen is attached to an actuator, called Pan-Tilt Unit, which controls the pose of screen, here only rotation around the vertical axis, which represents head motion typically appeared when a person turns his/her eyes to another participant, is used. The head pose captured by visual face tracking from videos and/or motion capture devices. Fig. 2 (a) shows an example of the time series of head pose angles, including measured values and the values for PTU control, which is a simplified version of actual head motion.

Fig. 2 (b) shows an example of the projection images sent to the projectors. They cover the face and bust of each person and are centered on the face. The projection images are extracted from the captured images with cameras in Fig. 1(a), and are transformed so that the skew-correct images appear on the screen regardless of the screen pose, as shown in Fig. 1(b) Right. The transformation is synchronized with the screen pose as in Fig. 2 (a). The projectors are calibrated at system installation.

4. FUTURE PERSPECTIVE

This demo paper presents our initial system as a platform for investigating how conversations can be more clearly delivered to remote viewers. A comprehensive survey, a complete description of the system, and evaluation results will be described in another conference paper. So far, although there has been a lot of research effort in the field of visual communication, authors believe that this work can provide a new driving force in terms of introducing the idea of meeting analysis, i.e. extracting people's behavior and using it for enhanced understanding. We are now working to extend our system towards a communication system that realizes multiparty-to-multiparty conversations across multi-locations.

5. REFERENCES