Observing Semantic Human Expression in a Group Meeting by using an Omni-directional Camera

Yusuke Soneda[1] Stirapongsasuti Sopicha[1] Yutaka Arakawa[1] Keiichi Yasumoto[1]

Abstract: In Japan, according to a survey of a company, average of meeting time per a day is 68.2 minutes. To improve the productivity of a group meeting, we need to know and measure the quality of the meeting. On the other hand, it is hard to record and analyze the conversation because of the reason of compliance. Therefore, we start considering the camera-based group meeting assessment that utilizes Omni-directional camera. We can observe various kind of facial expression and body gestures of attendees through the video. However, there is no standardized definition of human expressions. So, in this paper, we try to define the semantic human expression observed in a group meeting.

1. Introduction

A meeting is widely considered to be the most important key of achievement at work. Some people, especially office workers, have to participate at least one meeting in a week. However, some meetings have a low effectiveness due to various factors. For example, some meetings are hold so long time which can cause some participants lacking interest of the meeting contents. Also, participants' contribution in a meeting may be not equal because not all participants have the same chance to express their ideas. Therefore, it is necessary to develop a system to evaluate a productivity of a meeting. To fulfill this requirement, the type of meeting is the first issue which we should consider. Generally, there are 2 main types of meeting i.e. face-to-face meeting and online meeting.

In Japan, JR Tokai Agency [1] has surveyed about time and types of meeting hold in Japanese companies. The result shows that 91.5% of companies hold face-to-face meetings and the average of meeting time per a day is 68.2 minutes. Because of this result, our research focused on the face-to-face meeting when human's physical expression is the key point of a system development.

To clarify, we assumed that the quality of meeting can be measured by participants' physical expressions such as smiling, laughing, speaking, nodding and etc. According to Peter's study [2], nonverbal communication is important for social relationship and shows an emotion. For instance, nodding has an explicit meaning of positive feeling and makes people easily understand an agreement during conversation. Another assumption is if we develop a system to support participants having more contribution in a meeting, the meeting time will be decreased due to participants' brainstorms.

In this paper, we observed the semantic human expression using an omni-directional camera (360 degree camera).We setup an experiment of a meeting with 4 participants and applied the facial detection algorithm called OpenFace to get the facial position and mapped them to the labeled data.

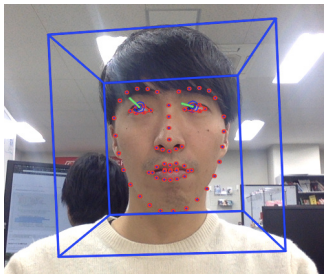

Fig. 1 OpenFace image

The paper is organized as follows. In the section 2 reviews the related work. Section 3 introduces the camera-based group meeting assessment. Section 4 shows how to label data. The last section is the conclusion of this paper.

2. Related Work

In order to evaluate a quality of meeting, some previous studies has been proposed. For example, Yu et al. [3] has proposed a smart meeting system using several cameras setting up in front of individual participants to record and detect who is talking to whom. They also constructed a meeting browser to provide other users understanding the meeting contents as a real time. Although the proposed method is interesting, but deployment is difficult to apply in the real scenario due to the number of cameras based on participants.

Recently, NTT Communication [4] has developed the system which can analyze the group meeting in real time. The proposed method used 8-channel microphone array with an omni-directional camera to record a meeting and analyze a meeting by speech recognition and visual processing. The result showed that NTT can achieve 80% of spoken words within 3 seconds. However, their system is suffered from high cost equipment. More-

[1] Nara Institute of Science and Technology, Ikoma, Nara 630–0101, Japan

© 2019 Information Processing Society of Japan

Fig. 2 Image of meeting

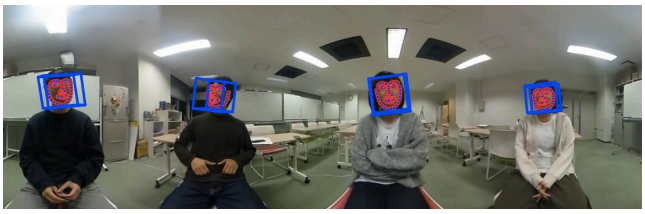

Fig. 3 OpenFace apply image for meeting

over, the main issue is how to deal with the privacy concern because microphones recorded contents of a meeting.

Obviously, designing such a system has a lot of issues to deliberate such as how to select commercial equipment, how to annotate participants' physical expression. To solves these issues, we outline the method in the following section.

3. Camera-based Group Meeting Assessment

3.1 Meeting recording system by using Omni-directional camera

We use a omni-directional camera, RICOH THETA S[5], and Macbook Pro to record the meeting. RICOH THETA S has two lens, so it can record 360 degree images. The resolution recorded by THETA S is 1920 × 1080 and number of frames for recording is 30 fps. For detecting the facial expression, we utilize OpenFace[6], Fig. 1 shows the example of output image.

3.2 Data collection

We assume that participants of meeting are 4 people, and meeting time is 10 min. They are same grade and know each other well. We put it with tripod and participants sitting around the omni-directional camera.

4. Labeling

We can analyze the faces of multiple persons simultaneously using omni-directional camera, we think that we can analyze not only the information from the face but also the influence of facial expressions in the future.

I described the facial expressions and attitudes identified in Table 1.

5. Conclusion

In this paper, we defined the labeling of the meeting to evaluate quality of meeting. We focused on the face movement so we used OpenFace.But body language is very useful for analyzing meeting too. Arakawa's research suggested some gestures show emotion[7].openpose[8] can analyze body movement like OpenFace.

When we finish making a model which can detect the important gesture in the meeting, we will apply the model real-time. If we show the gesture of the meeting real time, there is possibility of behavior change.

Table 1 definition of facial expressions

label	definition
smiling	Raise mouse corners.According to Peter's study[2] ,smiling is effective way to close interpersonal distance
laughing	Smiling with voice. We tried to predict smiling and laughing as a different one.
speaking	Say opinions or something. We do not use sound information to define speaking because the topic of meeting is privacy information. We used only mouse movement data to detect speaking.
nodding	Move head up and down. Nodding influences positive impression for speaker.
gazing	To look a participant. Eye contact expresses agreement for speaker or insists on listening to discussion.

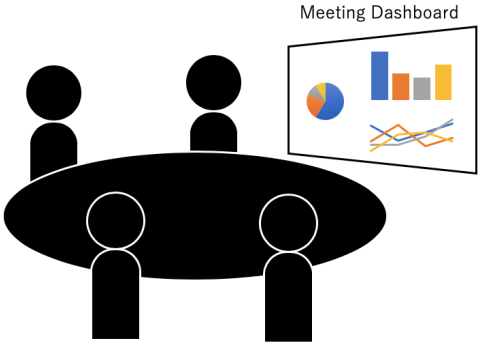

Fig. 4 Real time analyze meeting

Some participants said "I concerned omni-directional camera because the height of the camera is the same as the height of the line of sight".We will think about optimum equipment again.

References

[1] JR Tokai Agency. ビジネスパーソンの「社内会議」に関する調査. ビジネスパーソンウォッチング調査 *(in Japanese)*, 2017.
[2] Peter E Bull. *Posture and Gesture*. Pergamon Press, 1987.
[3] Zhiwen Yu, Motoyuki Ozeki, Yohsuke Fujii, and Yuichi Nakamura. Towards smart meeting: enabling technologies and a real-world application. In *ICMI*, 2007.
[4] T. Hori, S. Araki, T. Yoshioka, M. Fujimoto, S. Watanabe, T. Oba, A. Ogawa, K. Otsuka, D. Mikami, K. Kinoshita, T. Nakatani, A. Nakamura, and J. Yamato. Low-latency real-time meeting recognition and understanding using distant microphones and omni-directional camera. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(2):499–513, 2012.
[5] Ricoh theta s. https://theta360.com/ja/about/theta/s.html.
[6] Openface. https://github.com/TadasBaltrusaitis/OpenFace.
[7] Ayumu Arakawa. Naoto Suzuki. しぐさと感情の関係の探索的研究. *The Japanese Journal of Research on Emotion*, 10(2):56–64, 2004.
[8] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. In *arXiv preprint arXiv:1812.08008*, 2018.

© 2019 Information Processing Society of Japan