Augmenting Visual Cognitive Interactions: from Wearable First-person View to Ubiquitous Third-person Multi-views

Kenji Mase

Graduate School of Informatics, Nagoya University
E-mail: mase@nagoya-u.jp

1. Introduction

Human visual cognition and act play important roles in interactions with other entities such as human, things, events and intelligent environments. The visual cognition leads to recognition of objects and events as well as understanding of the behaviors and thoughts of other agents. The visual act, e.g. gaze, conveys interests and even emotions to others. Understanding the visual cognition and act will support interactions among people, things and intelligent environments. The exploitation of such interactions would realize a symbiotic society in the era of artificial intelligent assistive living, education, medicine, manufacturing and social networking in the near future.

This talk covers visual cognition and act research issues with several topics related to first-person view analysis by wearable sensing and the summarization of ubiquitously taken multi-view videos. The way of gazing and watching things contains information of personal interests in the view, which is a reflection of viewer's knowledge and skills about the things. Such knowledge and skill will be computationally formalized to be used in wisdom computing environment. The technical challenges are in (i) locating such parts-of-interest in the view over time and (ii) finding common interest against things among people and (iii) interpreting such behavior for further utilizations.

2. Ubiquitous Experience Media

First-person view video with a wearable camera captures user's experiencing event from his/her own view point without annoyance of manipulating camera toward the target of interest. With some clue of automatically annotating objects in the view, e.g. object recognition or IR tags, we can obtain more detail level of descriptions of user's interest at time. If all agents wear such devices, the events in the space can be described by sub-networks of visual act interactions, or interest relations. (see figure 1)

Social network and/or enthusiastic event in the space is extractable, if desired. Personal experience is annotated by a corrective intelligence. Privacy issue may arise for such deep interactions of participants and holistic capture of events in general public places. However, such practices will be appreciated for safety and efficiency such as in hospitals, elderly care-houses and some factories.

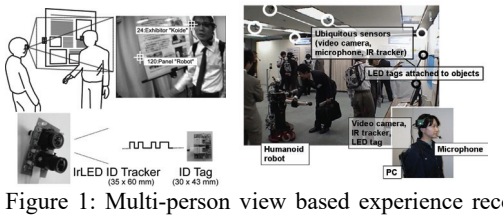

Figure 1: Multi-person view based experience recording and sharing with IrLED tags [1]

3. Multi-view Video Recommendation

When such videos are taken by other people, i.e. third person, the video contents become subject of summarization based on the viewing people's interests. We had recorded soccer games with about 20 fixed cameras. Viewers are given a large freedom of viewing position selection, but with a cumbersome selection task that would destroy the entertainment of watching game, or barrier for coaching.

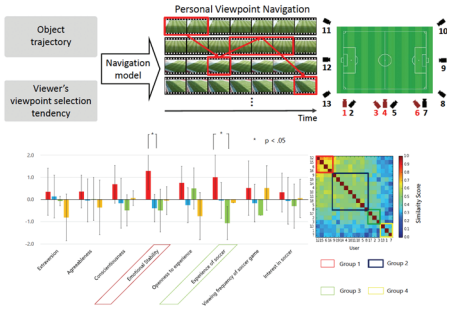

Figure 2: Multi-view video content view recommendation algorithm gives better choice for the expert and motivated group [4].

We have therefore developed an automatic viewpoint recommendation algorithm that uses positions of players and a ball as general inputs and the editing example as the machine learning reference [2]. The resulting recommendation scheme is best fit for active and expert viewers of the game while the performance is low for lazy viewers (see figure 2).

4. Gaze Analysis of Expertise

Gaze is a proactive display of personal interests. The eye tracking device becomes available in portable and wearable forms with less constraint usability. The figure 3 example shows gaze position difference between a soccer coach expert and novices. Coaching skill is easily observed from the gaze distribution pattern [3].

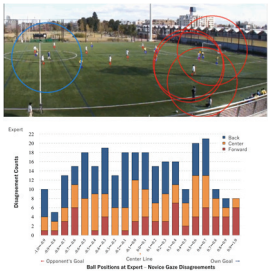

Figure 3: Eye gaze point and distribution along soccer field. Expert coach (blue) doesn't follow ball traces, but checks players' readiness for next moves.

References

[1] Kenji Mase, et al. "Ubiquitous Experience Media", IEEE Multimedia, pp.20-29, Oct-Dec. 2006.
[2] Xueting Wang, et al. "Viewpoint Sequence Recommendation Method based on Contextual Information for Multi-view Video", IEEE Multimedia, pp. 40-50, Oct.-Dec. 2015.
[3] Atsushi Iwatsuki, et al. "Skilled Gaze Behaviors Extraction Based on Dependency Analysis of Gaze Patterns on Video Scenes", ACM ETRA 2016, 2016.3.
[4] Xueting Wang, PhD thesis. Nagoya Univ. 2018.3