Challenges in Gaze-based Intention Recognition

Author(s):
Kiefer, Peter

Publication Date:
2018

Permanent Link:
https://doi.org/10.3929/ethz-b-000311201

Rights / License:
Creative Commons Attribution 3.0 Unported
5.16 Challenges in Gaze-based Intention Recognition

Peter Kiefer (ETH Zürich, CH)

License Creative Commons BY 3.0 Unported license © Peter Kiefer

5.16.1 Introduction

Recent technological developments have made ubiquitous gaze sensing a goal realistically reachable in the near future. This enables new generations of systems that adapt to the user’s gaze in order to provide assistance. It is expected that intelligent assistance can be provided by recognizing cognitive states of a user [1]. While previous work has considered the gaze-based recognition of cognitive states such as interest [2], boredom [3], or cognitive load [4], this extended abstract discusses the recognition of a cognitive state that is on a particularly high cognitive level: intention.

5.16.2 Challenges

The following challenges for gaze-based intention recognition can be identified:

Establishing a common terminology and framework: the terminology related to cognition is not used consistently throughout the Human Computer Interaction (e.g., [5]) and eye tracking literature (e.g., [6]). This includes terms, such as, cognitive state, intention, plan, activity, action, cognitive load, goal and task. There is a need to review previous work in these and related fields to establish a common ground.
Selection of gaze features for building the models: different features of gaze have been suggested for gaze-based activity recognition (e.g., [7]). These need to be considered, combined, and possibly extended for inferring higher-level cognitive states.

Bringing together short-term and long-term models: the models for short-term prediction (i.e., in the range of seconds or milliseconds, e.g., [8]) established in the eye tracking and vision research communities need to be combined with models for longer term intention recognition and prediction (i.e., in the range of several minutes) well-known in Artificial Intelligence and Cognitive Science (e.g., [9, 10]).

Accounting for hierarchical, parallel and interleaved intentions: intentions can be seen as hierarchical concepts (i.e., an intention can be implemented by several sub-intentions) that occur in parallel (i.e., a subject may have several intentions at the same time) or interleaved (i.e., an intention can be ‘paused’ and superseded by some other intention for a while, but picked up again later) (e.g., refer to [11]).

Bottom-up vs. top-down: there is a need to re-visit classic discussions in the literature regarding the benefits, drawbacks and potential combinations of data-driven and model-driven approaches.

Computational methods and platforms for gaining efficiency: gaze data come at high frequency, and the acceptable time lag between the occurrence of an intention and the according assistance is small. This calls for efficient algorithms and computing platforms.

Context-awareness: adding context to the inference model will benefit the recognition accuracy. In particular, knowing the current situation of the user (such as, being at work or in a restaurant) will help in disambiguating which kinds of intentions are possible in that situation (e.g., [12]).

Relation to affective states: the relation between affective states (possibly also inferred from eye movements) and intentions requires further investigation.

Research practices and infrastructure: one challenge for the community consists in creating and sharing gaze datasets for different domains, annotated with intentions, which can be used for benchmarking purposes.

References


