Chapter 74
Inferring Intent and Action from Gaze in Naturalistic Behavior: A Review

Kristian Lukander
Finnish Institute of Occupational Health, Helsinki, Finland

Miika Toivanen
Finnish Institute of Occupational Health, Helsinki, Finland & Department of Teacher Education, University of Helsinki, Helsinki, Finland

Kai Puolamäki
Finnish Institute of Occupational Health, Helsinki, Finland

ABSTRACT

We constantly move our gaze to gather acute visual information from our environment. Conversely, as originally shown by Yarbus in his seminal work, the elicited gaze patterns hold information over our changing attentional focus while performing a task. Recently, the proliferation of machine learning algorithms has allowed the research community to test the idea of inferring, or even predicting action and intent from gaze behaviour. The on-going miniaturization of gaze tracking technologies toward pervasive wearable solutions allows studying inference also in everyday activities outside research laboratories. This paper scopes the emerging field and reviews studies focusing on the inference of intent and action in naturalistic behaviour. While the task-specific nature of gaze behavior, and the variability in naturalistic setups present challenges, gaze-based inference holds a clear promise for machine-based understanding of human intent and future interactive solutions.

DOI: 10.4018/978-1-7998-2460-2.ch074

INTRODUCTION

Gaze tracking in psychological, cognitive, and user interaction studies has recently evolved toward mobile solutions, which enable direct assessment of users' visual attention in natural environments. The capability for reliably tracking users' locus of attention with wearable devices has developed quickly as the device manufacturers have miniaturized their technology to wearable eye-glass-like frames, with a number of open-source solutions adding their contribution to the variety[1]. Also, increases in signal processing power and recent developments in gaze tracking algorithms now enable complex tracking methods to operate in portable devices, even in real-time (Toivanen et al., 2017).

Human eye movements shift the focus of attention to gather visual information for action planning. Conversely, they can be used to provide information for inferring users' intentions and next actions. However, gaze behavior in natural, unstructured tasks is markedly complex. Models created in controlled laboratory environments do not often satisfactorily explain such natural gaze behavior. While laboratory studies in gaze tracking typically aim for isolating single components of behavior to accurately model and study some part of the human visual system or cognition, natural gaze behavior involves a complex interplay of these cognitive processes. The modeling of these processes computationally is difficult, not least because of the unknowns involved: it is a challenge to construct an experimental setup with a known "ground truth" for training, e.g., a machine learning model. In addition, the methods and implementations of machine learning applied to gaze data are still often customized and fine-tuned for each task at hand. This results in a set of isolated, individual contributions to gaze-based inference which are slowly converging to a more generic understanding on gaze-action behavior.

The issue of inferring user action with mobile gaze tracking is highly multidisciplinary, requiring deep understanding of a variety of research fields. These include the functioning of human visual system, mathematical modeling, computer vision, machine learning, cognitive processes, user interaction, and psychology. Here, we review current advances in attempting to infer the cognitive task of users based on their gaze behavior.

BACKGROUND

Motivation

Work toward this paper started from organizing the workshop[2] on "Inferring user action with mobile gaze tracking" as part of the Mobile HCI 2016 conference in Florence, Italy (Toivanen et al., 2016). The objective of the workshop was to map out the developing field of task and intent recognition in natural gaze interaction. The round-up talk after the workshop forms the basis of this contribution.

Eye and Gaze

The human visual system constantly samples the environment through a spatial window, where – due to the distribution of photoreceptor cells on the retina – high acuity information can only be obtained from the central area of the fovea, spanning about 1.5 degrees of visual angle. While the percept we experience seems stable, we inspect the scene through a constant stream of rapid, ballistic eye movements, saccades, to acquire new features from within the visual field. The acquisition of information