3D Gaze Estimation from 2D Pupil Positions on Monocular Head-Mounted Eye Trackers

Mohsen Mansouryar  Julian Steil  Yusuke Sugano  Andreas Bulling
Perceptual User Interfaces Group
Max Planck Institute for Informatics, Saarbrücken, Germany
{mohsen, jsteil, sugano, bulling}@mpi-inf.mpg.de

Figure 1: Illustration of the (a) 2D-to-2D, (b) 3D-to-3D, and (c) 2D-to-3D mapping approaches. 3D gaze estimation in wearable settings is a task of inferring 3D gaze vectors in the scene camera coordinate system.

Abstract

3D gaze information is important for scene-centric attention analysis, but accurate estimation and analysis of 3D gaze in real-world environments remains challenging. We present a novel 3D gaze estimation method for monocular head-mounted eye trackers. In contrast to previous work, our method does not aim to infer 3D eyeball poses, but directly maps 2D pupil positions to 3D gaze directions in scene camera coordinate space. We first provide a detailed discussion of the 3D gaze estimation task and summarize different methods, including our own. We then evaluate the performance of different 3D gaze estimation approaches using both simulated and real data. Through experimental validation, we demonstrate the effectiveness of our method in reducing parallax error, and we identify research challenges for the design of 3D calibration procedures.

Keywords: Head-mounted eye tracking; 3D gaze estimation; Parallax error

1 Introduction

Research on head-mounted eye tracking has traditionally focused on estimating gaze in screen coordinate space, e.g., of a public display. Estimating gaze in scene or world coordinates enables gaze analysis on 3D objects and scenes and has the potential for new applications, such as real-world attention analysis [Bulling 2016]. This approach requires two key components: 3D scene reconstruction and 3D gaze estimation.

In prior work, 3D gaze estimation was approximately addressed as a projection from estimated 2D gaze positions in the scene camera image to the corresponding 3D scene [Munn and Pelz 2008; Takemura et al. 2014; Pfeiffer and Renner 2014]. However, without proper 3D gaze estimation, gaze mapping suffers from parallax error caused by the offset between the scene camera origin and eyeball position [Mardanbegi and Hansen 2012; Duchowski et al. 2014]. To fully utilize the 3D scene information it is essential to estimate 3D gaze vectors in the scene coordinate system.

While 3D gaze estimation has been widely studied in remote gaze estimation, there have been very few studies in head-mounted eye tracking. This is mainly because 3D gaze estimation typically requires model-based approaches with special hardware, such as multiple IR light sources and/or stereo cameras [Beyer and Flickner 2003; Nagamatsu et al. 2010]. Hence, it remains unclear whether 3D gaze estimation can be done properly only with a lightweight head-mounted eye tracker. Świrski and Dodgson proposed a method to recover 3D eyeball poses from a monocular eye camera [Swirski and Dodgson 2013]. While it can be applied to lightweight mobile eye trackers, their method has been only evaluated with synthetic eye images, and its realistic performance including the eye-to-scene camera mapping accuracy has never been quantified.

We present a novel 3D gaze estimation method for monocular head-mounted eye trackers. Contrary to existing approaches, we formulate 3D gaze estimation as a direct mapping task from 2D pupil positions in the eye camera image to 3D gaze directions in the scene camera. Therefore, for the calibration we collect the 2D pupil positions as well as 3D target points, and finally minimize the distance between the 3D targets and the estimated gaze rays.

The contributions of this work are threefold. First, we summarize and analyze different 3D gaze estimation approaches for a head-mounted setup. We discuss potential error sources and technical difficulties in these approaches, and provide clear guidelines for designing lightweight 3D gaze estimation systems. Second, following from this discussion, we propose a novel 3D gaze estimation method. Our method directly maps 2D pupil positions in the eye camera to 3D gaze directions, and does not require 3D observation from the eye camera. Third, we provide a detailed comparison of our method with state-of-the-art methods in terms of 3D gaze esti-