Hybrid model and appearance based eye tracking with Kinect

Kang Wang*
Rensselaer Polytechnic Institute

Qiang Ji†
Rensselaer Polytechnic Institute

Abstract

Existing gaze estimation methods rely mainly on 3D eye model or 2D eye appearance. While both methods have validated their effectiveness in various fields and applications, they are still limited in practice, such as portable and non-intrusive system and robust eye gaze tracking in different environments. To this end, we investigate on combining eye model with eye appearance to perform gaze estimation and eye gaze tracking. Specifically, unlike traditional 3D model based methods which rely on cornea reflections, we plan to retrieve 3D information from depth sensor (Eg, Kinect). Kinect integrates camera sensor and IR illuminations into one single device, thus enable more flexible system settings. We further propose to utilize appearance information to help the basic model based methods. Appearance information can help better detection of gaze related features (Eg, pupil center). Plus, eye model and eye appearance can benefit each other to enable robust and accurate gaze estimation.

Keywords: 3D eye model, eye appearance, learning, depth sensor

Concepts: •**Computing methodologies** → **Computer vision; Machine learning;**

1 Research Objective

3D eye model are mainly used in model based gaze estimation methods, which are known for the ability to handle free head movement. Among them, IR lights based methods can achieve high gaze estimation accuracy. But the hardware setup is rather complex including infrared lights, filters, etc. An alternative for model based methods is to use depth sensor. System settings are simplified by integrating all the hardware into one single device, while keep the ability for pose-free eye gaze tracking. Eye appearance are typically utilized in appearance based methods, which consider the gaze estimation problem from learning point of view, by mapping 2D eye appearance to 2D gaze positions on the display surface. While appearance based methods can achieve minimum hardware (one web camera), they typically suffer from head movement issues, unless an extra correction model is built to compensate the appearance change resulted from pose variations. Besides, appearance based methods may not be applied to applications where gaze directions are required or no specific display surface is available.

Our goal is to make better use of 3D eye model and eye appearance. We plan to build a hybrid model to further enhance gaze tracking experience. Specifically, we have following objectives:

- minimum hardware with only one depth sensor: Kinect.
- can work in less constrained settings: different head poses, illuminations(Kinect can still work probably), no explicit personal calibration required, etc.
- achieve real time gaze tracking with reasonable gaze estimation accuracy.

2 Problem Statement

To achieve the objective, we need to solve the following problems step by step:

- First, we plan to build a basic model based system with Kinect. This system is able to perform gaze estimation in typical environments, achieve reasonable accuracy with personal calibration.
- Second, we plan to fulfill following goals to enhance the system:
  - ⋆ Explore more on eye appearance information to improve iris/pupil center detection. Challenging environments significantly degrade the detection performance of iris/pupil center, therefore we plan to better utilize eye appearance information to achieve robust and accurate detection.
  - ⋆ System optimization to achieve real time gaze tracking.
  - ⋆ Achieve the goal of implicit personal calibration.
- Third, we plan to build a hybrid model to better combine eye model and eye appearance. Beyond simply using appearance information to help feature detection in eye model as in Step 2, we try to systematically explore the underlying relationship between eye model and eye appearance. A proper hybrid model can be constructed based on the relationship and help gaze tracking experience and possible eye related applications.

3 Approach and method

3.1 Model based gaze estimation with Kinect

By analyzing human eye anatomy, we build the 3D eye model as shown in Fig. 1. Gaze directions is defined as the visual axis that passes through fovea and cornea center. Gaze direction can be estimated given the head pose information $\{\mathbf{R}, \mathbf{T}\}$, 3D pupil center $\mathbf{p}$ and related personal parameters. The gaze estimation procedure is similar to the work by [Sun et al. 2014]. However, our 3D eye model use more personal parameters by considering the offset between optical and visual axis and that visual axis passes through cornea center.

3.2 System enhancement

Pupil/iris detection may face with following error factors that significantly degrade the detection accuracy: 1) challenging environments including various illuminations, and low image resolutions; 2) various occlusions including eyelashes, eyelids, hair,

*e-mail: wangk10@rpi.edu
†e-mail: qji@ecse.rpi.edu

ETRA '16, March 14-17, 2016, Charleston, SC, USA
ISBN: 978-1-4503-4125-7/16/03
DOI: http://dx.doi.org/10.1145/2857491.2888591

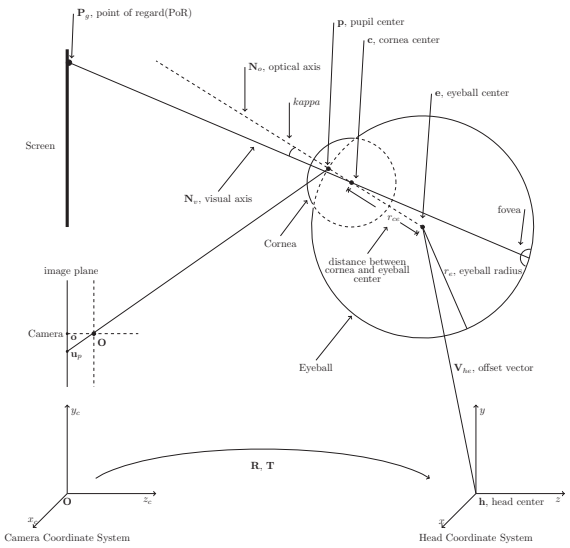

Figure 1: *3D eye model.*

reflections, etc; 3) different head poses. Recent work on pupil/iris detection mentioned in the survey paper [Song et al. 2013] mainly focuses on traditional appearance or shape features, but these features might not yield good results under challenging settings. Therefore, we plan to leverage on deep network's [Krizhevsky et al. 2012] representation power to learn better feature representations of pupil/iris. Specifically, we treat pupil/iris center detection as an binary object classification problem. To this end, we plan to use patch-based method to generate positive patches which include iris region and negative patches which include other unrelated regions. Finally, we can obtain better feature representations of iris which help improve pupil/iris center detection.

To eliminate explicit personal calibration, we plan to follow the visual saliency based method similar to [Y. Sugano, Y. Matsushita, and Y. Sato 2013]. But we need to propose a new model to adapt to gaze tracking systems with depth sensor.

To enable real time gaze tracking, we need to optimize different components of the system. Iris/pupil center detection is the most time-consuming step, we plan to accelerate the detection algorithm. Specifically, we plan to treat pupil/iris center as one facial landmark, and use the supervised descent method [Xiong and De la Torre 2013] to detect the pupil/iris center given the learned features from deep networks.

3.3 Hybrid model by combining 3D eye model and eye appearance

3D eye model encodes the eye anatomy knowledge, while eye appearance are typically used in learning based methods. Several attempts have been made to incorporate knowledge into learning. First, traditional appearance based methods only use eye appearance information in the learning framework. Given the knowledge that gaze direction is determined by head pose and eye pose, several methods add head pose information to alleviate the head movement issues. The head pose information can be estimated by 2D facial landmarks, stereo vision systems or depth sensors (Kinect). These methods, however, only establish a superficial relationship between eye appearance and eye model and may not give good results. In [Funes Mora and Odobez 2014], the authors utilize the eye model knowledge by projecting the 3D eye model to 2D eye appearance, such relationship are encoded into a generative graphical model, and a head pose rectified eye image can be inferred from the model.

The approach in [Funes Mora and Odobez 2014] offers us a good starting point to combine eye model with eye appearance. But we want to explore more by analyzing the generation process of eye gaze and how eye gaze related to eye appearance. We also plan to use a graphical model to relate eyeball center, pupil center, eye pose, head pose, eye appearance, etc. Given the model, we can utilize both eye anatomy knowledge and the eye appearance information to infer eye gaze, or vice versa.

4 Results to Date

For Section 3.1, we have implemented the basic model based gaze estimation methods. The average gaze estimation error for different subjects and different head poses is approximately 4 degree.

For Section 3.2, we are working on building a simple deep network to perform pupil detection. Besides, we have implemented the supervised descent method with SIFT features to detect pupil center in real time. Ignoring the case where face/eye is not detected, the overall detection accuracy is comparable with state of the art methods on BioID dataset [BioID]. But the algorithm has not been tested on other datasets with much more challenging cases. In addition, we have proposed an implicit personal calibration method with visual saliency, and is validated on a gaze tracking system with infrared lights (high accuracy). We are working on extending the implicit calibration algorithm on Kinect based systems.

5 Plans for future work

We have already built the basic model based gaze tracking system. Besides further testing and improvement on the basic system, our future work lies on: 1) learn representative features of pupil/iris through deep networks; 2) apply the implicit calibration algorithm to the basic system, with proper modifications; 3) work on the hybrid models, begin with determining the graphical model structures to relate eye model and eye appearance.

References

BIOID. Bioid database https://www.bioid.com/about/bioid-face-database.

FUNES MORA, K. A., AND ODOBEZ, J. 2014. Geometric generative gaze estimation (g3e) for remote rgb-d cameras. *CVPR*.

KRIZHEVSKY, A., SUTSKEVER, I., AND HINTON, G. E. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*.

SONG, F., TAN, X., CHEN, S., AND ZHOU, Z.-H. 2013. A literature survey on robust and efficient eye localization in real-life scenarios. *Pattern Recognition*.

SUN, L., SONG, M., LIU, Z., AND SUN, M. 2014. Real-time gaze estimation with online calibration. *IEEE Multimedia and Expo*.

XIONG, X., AND DE LA TORRE, F. 2013. Supervised descent method and its application to face alignment. *CVPR*.

Y. SUGANO, Y. MATSUSHITA,, AND Y. SATO. 2013. Appearance-based gaze estimation using visual saliency. *IEEE Trans. Pattern Analysis and Machine Intelligence*.