3D Gaze Tracking and Automatic Gaze Coding from RGB-D Cameras

Kenneth Alberto Funes Mora Jean-Marc Odobez
Idiap Research Institute, CH-1920, Martigny, Switzerland
École Polytechnique Fédéral de Lausanne, CH-1015, Lausanne, Switzerland
{kfunes, odobez}@idiap.ch

Abstract

Gaze is recognized as one of the most important cues for the analysis of the cognitive behaviors of a person such as the attention displayed towards objects or people, their interactions, functionality and causality patterns. In this short paper, we present our investigations towards the development of 3D gaze sensing solutions from consumer RGB-D sensors, including their use for the inference of visual attention in natural dyadic interactions and the resources we have made or will make available to the community.

1. Introduction

For a machine to understand how does a human interact with the environment and people, it requires to sense diverse cues describing the physical state of the individual, like head and body pose, hands position, gestures, facial expressions or gaze. These cues, in conjunction with a description or the sensing of the environment, provide the building blocks necessary to infer higher cognitive information about the mental state of the person, such as emotions, intentions, objects functionality or causality reasoning.

Amongst the non-verbal cues, gaze is recognized as one of the most important one. Many computer vision based techniques have been proposed over the last 30 years to estimate it [7]. Current systems, however, are either highly expensive and based on specialized infrared (IR) hardware [6], or based on less costly cameras but depending on methods restraining their utility e.g. in case of user mobility.

Recently, the advent of cheap RGB-D sensors offers new opportunities for gaze sensing and analysis. Thanks to the depth information, both 3D scene modeling as well as head pose estimation and eye localization have become simpler, allowing to analyze a subject's attention by computing the intersection of the 3D direction of his gaze estimate with scene surface elements. In the following we describe our endeavors towards the development of a remote (not head mounted) 3D gaze estimator based on consumer RGB-D sensors and its use for gaze inference in natural interactions.

2. RGB-D based 3D gaze estimation

RGB-D based sensing allows for 3D scene modeling, with correct texture-shape binding and no scale ambiguity (provided calibrated sensors). We exploited these characteristics to develop gaze estimation strategies as follows.

2.1. Head pose free gaze estimation

Appearance based methods (ABM) have gained important attention recently. By modeling directly the mapping from the eye image to the low-dimensional gaze parameters space, they avoid the difficult local features tracking (such as the iris), making them suitable for the low-resolution sensing conditions that commonly result from the use of consumer sensors or in less restraining scenarios.

However, ABM have important limitations. In particular, by requiring test data to be very similar to the training data, they need either large training sets, or session-specific training sessions. As a corollary, they are sensitive to different users, ambient sensing conditions, and importantly, to variations in the head pose, even for a single user.

To address gaze estimation under free head pose, we proposed an effective approach [3] that leveraged on the RGB-D multimodality: depth data has proven to be valuable for accurate head pose tracking, while standard imaging is needed to infer the gaze from the eye image.

The method first consists of rectifying the eye image by combining head pose and depth information so that the eye appears as always seen from a single (frontal) pose. Such rectification is adequate for ABM, as it reduces the amount of needed training data. We have validated this approach in [3] using a sparse reconstruction gaze estimation technique [8]. Furthermore, in [4] we extended the ABM approach to handle gaze inference for unseen people and directly from both eyes rather than independently from each eye.

2.2. Geometric generative gaze estimation

ABM are appealing but mix individual eye geometry with eye appearance and ambient illumination, making it difficult to perform adaptation to any of these elements.

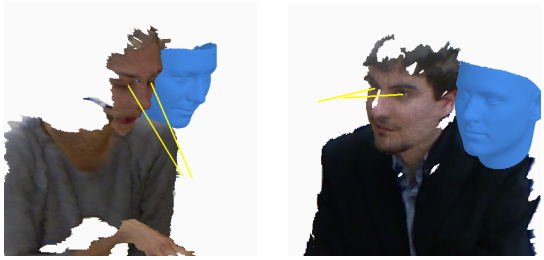

Figure 1. Dyadic interaction (each person is rendered separately). Displaying the estimated head pose and gaze.

To address this issue, we recently proposed a method which model the head-pose rectified eye images as an appearance generative process relying on the eyeball geometry [5]. The problem is then formulated in terms of a global eye-image likelihood, constrained by the eyeball geometry (and gaze orientation). As a result, the latent geometry can be inferred by maximizing a global measure, rather than by fitting local features as commonly done in classical geometric based methods (reason why these methods require high resolution and high contrast data) [7].

We have called this methodology geometric generative gaze estimation (G^3E) and it has several advantages with respect to previous proposals. It is suitable for low resolution imaging and it decouples ambient conditions from the user specific geometry (using a semantic segmentation), making it capable for adaptation to conditions different to those of the training data. The model's geometric prior makes it appropriate for training from a few samples and extrapolating to other conditions. This method is also head-pose invariant as it is based on the framework we developed (cf. Sec. 2.1).

2.3. Automatic gaze coding

The proposed methods can be used to remotely sense the 3D gaze direction, and we exploited this ability in the case of natural dyadic interactions consisting of job interviews, as illustrated in Fig. 1. This is a valuable application, where gaze, along with other cues, can be used to infer personality traits and hireability of the subjects, and more generally can help to clarify a spoken message, and to decode the subject's intentions or emotions. On the other hand, manual annotations are difficult and highly time consuming.

In this scenario, detecting whether a person is looking at another one becomes a geometric problem and can be formulated as the detection of when the 3D gaze direction intersects the head position of the other subject. Using a dual RGB-D sensor pair pointed to each subject, we proposed a simple calibration method and applied our ABM generic gaze estimation framework to each individual [4].

Despite the subtlety of the gazing and eyelid patterns, our method showed to be highly accurate in terms of frame-wise gazing events detection (85%) and significantly improved over head pose only visual attention estimation (65%) [2]. This validates the potential of our overall methodology.

3. Resources

EYEDIAP dataset [1] In spite of the importance of gaze estimation, almost no bechmark data is available[2].Therefore, we have recently released the EYEDIAP public database [1] involving 14 people and comprising RGB-D and high resolution videos with associated ground truth.

The recording methodology was designed to systematically include, and isolate, most variables which affect the remote gaze estimation algorithms: i) Head pose variations; ii) Person variation; iii) Changes in ambient and sensing conditions and iv) Types of target: screen or 3D object.

Gaze estimation code. We will release our code for head pose tracking and free gaze estimation for ABM [3]. To this end, we also provide tools for fitting a deformable mesh model to specific individuals and to train gaze appearance models (as an alternative to generic models [4]).

We believe this will provide useful tools for researchers working in HRI or social interaction analysis.

Acknowledgments Authors gratefully acknowledge the support from the Swiss National Science Foundation (Projects: 200020_153085, G3E and CRSII2_147611, UBImpressed) www.snf.ch.

References

[1] K. A. Funes Mora, F. Monay, and J.-M. Odobez. EYEDIAP: A Database for the Development and Evaluation of Gaze Estimation Algorithms from RGB and RGB-D Cameras. In *Eye Tracking Research and Applications*, Safety Harbor, FL, 2014.

[2] K. A. Funes Mora, L. S. Nguyen, D. Gatica-Perez, and J.-M. Odobez. A Semi-Automated System for Accurate Gaze Coding in Natural Dyadic Interactions. In *Int Conf. on Multimodal Interaction*, Sydney, Dec. 2013.

[3] K. A. Funes Mora and J.-M. Odobez. Gaze estimation from multimodal Kinect data. In *Computer Vision and Pattern Recognition Workshops*, pages 25–30, June 2012.

[4] K. A. Funes Mora and J.-M. Odobez. Person Independent 3D Gaze Estimation From Remote RGB-D Cameras. In *International Conference on Image Processing*, Sept. 2013.

[5] K. A. Funes Mora and J.-M. Odobez. Geometric generative gaze estimation (G3E) for remote RGB-D cameras. In *Computer Vision and Pattern Recognition*, Ohio, June 2014.

[6] E. D. Guestrin and M. Eizenman. General theory of remote gaze estimation using the pupil center and corneal reflections. *Trans. on bio-medical engineering*, June 2006.

[7] D. W. Hansen and Q. Ji. In the eye of the beholder: a survey of models for eyes and gaze. *IEEE trans. on pattern analysis and machine intelligence*, 32(3):478–500, Mar. 2010.

[8] F. Lu, Y. Sugano, T. Okabe, and Y. Sato. Inferring human gaze from appearance via adaptive linear regression. In *Int. Conf. on Computer Vision*, Barcelona, Nov. 2011.

[9] B. A. Smith, Q. Yin, S. K. Feiner, and S. K. Nayar. Gaze Locking: Passive Eye Contact Detection for Human-object Interaction. In *Symposium on User Interface Software and Technology*, UIST '13, New York, NY, USA, 2013. ACM.

[1] www.idiap.ch/dataset/eyediap

[2] A recent exception is the CAVE dataset [9].