EyeCloud: Cloud Computing for Pervasive Eye-Tracking

Hana Vrzakova
School of Computing
University of Eastern Finland

Roman Bednarik
School of Computing
University of Eastern Finland

In order to move away from limited laboratory studies to pervasive mobile contexts, eye-tracking needs to be combined with technologies that allow instant access to large amounts of data of a massive number of users simultaneously. Cloud-computing seems to offer the solutions for these needs. In this conceptual paper, we map the forming research area from an interactive technology development point of view.
Keywords: pervasive, eye-tracking, implicit gaze data, cloud computing

Introduction

When dreaming about a perfect wearable interactive technology, users often pick up on mutually unattainable requirements. Low costs, low energy demands, accurate and reliable, engaging, and, of course, always ready to use, to name a few. The requirements and expectations grow with every advancement achieved. To satisfy such greedy markets, industry and academia expand their arsenal of technologies, and instead of mastering a single technology, they increasingly combine the best-of-breeds solutions. The wave of the demanding user, we believe, is coming to the beach of eye-tracking too, and surviving it requires bridging eye-tracking with other fast growing domains. This is the call for pervasive eye-tracking to marry cloud computing.

One minute of 120Hz eye-tracking recording typically includes a compressed HD video stream (file size of 10-15 megabytes) and associated gaze-data in a text format (file size of 1 megabyte). One day of eye-tracking recordings would take as much as tens of gigabytes. If one envisions a truly pervasive deployment of eye-tracking systems, such as a merge of eye-trackers and augmented reality, visual attention data cannot be stored and processed locally, for several important reasons on top of the unbounded storage space.

First and foremost, future interactive environments and systems will be ubiquitous and make *implicit* use of gaze data of a large number of users at once. For example, a shopping window will adapt to the passing customers depending on their interest and will recommend products based on patterns of other customers. An interactive TV will be actively collecting visual attention data from the distributed viewers to improve user experience. A traffic information system will detect and report lapses of attention and alertness of drivers on various road segments. A massive online course management system will monitor the attention and behaviors of the participants at once.

The common denominator in all of the examples is the requirement for the data of a massive number of distributed mobile users to be available for effective modeling and prediction of behaviors, contexts, and needs of single users.

The second evidence is the expected merge of eye-tracking and augmented reality systems. Plentiful calls exist for inclusion of gaze tracking as a source of adaptivity for augmented reality (Van Krevelen & Poelman, 2010). Industry increases the offer of advanced augmented reality systems, such as Google Project Glasses. At the incoming junction, massive online multimodal data will be linked with gaze. However, current eye-tracking solutions are not prepared for such on-the-fly storage and real-time processing demands.

Although the range of applications for implicit gaze-interaction is wide, let us envision a more concrete scenario in an online lecture, where spatial and temporal distribution of students plays a significant role. To achieve an equal classroom interaction, a teacher experiences hard times to cue students' attention through complex explanations, e.g. in math formulas or software comprehension. Such pedagogical task is challenging because smart pointing, especially using a mouse cursor, is far from natural.

Students, on the other hand, need to follow teacher's attentional cues to comprehend the materials. Yet, they do not have access to the attention of the teacher all the time and may fail to maintain focus during explanations. Furthermore, if the mass of students demands attention at the same time, the teacher will fail in management of the scarce resource.

Delayed understanding and comprehension, lost of focus and lack of attention decrease students' and teacher's engagement and thus lead to suboptimal educational performance. Here pervasive eye-tracking is a plausible solution, however, only if delivered unobtrusively and in real-time through a cloud.

Figure 1. Pervasive eye-tracking in online streaming lectures. In this case, shared visual cues of connected students and teachers can help to follow ongoing discussions. Adopted from (Lee Green, 2012).

If the concept of *EyeCloud* is applied in the aforementioned case, students and teacher enhance their lectures using *augmented reality eye-trackers* and interconnect them to the *cloud*. Both way, EyeCloud offers implicit and explicit information and interaction, datamined from the individuals' and group visual attention.

In the explicit use of gaze, students can see the teacher's points of interest and indicators of the topic importance, as illustrated in Figure 1, and clearly see when, where and to what to pay attention. Using the explicit gaze information, students can better allocate their resources and adapt learning strategies.

On the other hand, the teacher cannot benefit much from explicit cues since gaze points of dozens students are too detailed and thus overwhelming. Here implicit information, gained from intelligent predictions from gaze (Bednarik, Eivazi, & Hradis, 2012), can introduce cognitive-state overviews, such as students' engagement or fatigue. The teacher can thus estimate more easily when he should pause and ask for questions, even in the virtual classroom.

In this position paper, we open a discussion on expectations and requirements, and progress towards implicit pervasive eye-tracking. We believe the ubiquity of eye tracking creates an *unprecedented shift*. It is of great importance for eye-tracking community to prepare for the arrival of the new technology. We draft a research agenda for the EyeCloud: a cloud-based eye-tracking data storage, processing and applications.

The EyeCloud Agenda

The questions in the agenda of *EyeCloud* are many, but because our background lies in interactive technology development, we primarily address the issues pertinent to our field as we draw on experience with developing interactive eye-tracking technologies. Here we thus open the following questions and we invite the community to expand the list:

- Understanding of essential information for interaction
- Local vs. distributed storage and processing
- Interaction design of pervasive eye-tracking

Not all interactive eye-tracking applications will, however, require cloud computing interfaces. For example, gaze-control applications will perhaps not be coping with voluminous data. However, here we consider those uses of eye-tracking streams that will improve the user experience in more implicit ways than for an explicit interaction purposes. Because implicit gaze-data require access to shared databases and models, we argue that the need for efficient data storage, processing and modelling can be satisfied by cloud-based solutions. Our motivation for this conceptual work is to raise a discussion about limitations, benefits, potentials, and requirements for taking eye-tracking to the wild.

Interaction and Gaze Essentials

Every interaction design starts with question *What information is important to user?*, followed by *What is a natural behavior in interacting with the information?*. In case of EyeCloud, these two questions are central since their answers comprise visual attention, user reactions, application and cloud design, and finally overall user experience. In case of pervasive eye-tracking, these two questions become, however, tricky.

Interaction design in pervasive eye-tracking is tightly connected to the constrains and requirements specific to each user, context, employed eye-tracker, implemented application, cloud solution, and feasible infrastructure. All together they present a complex design space and one cannot require all variables to be satisfied at once. We argue that eye-tracking research can contribute to this call by *deeper understanding* of the essentials needed for pervasive eye-tracking.

Gaze essentials depend on eye-tracker limitations, user capabilities, and the understanding of what part of gaze-behavior provides most important information. A gaze essential can be one specific fixation, a gaze gesture, a complex fixation pattern or a blink. These and only eye movements provide relevant description of user gaze behavior and carry the links to the interaction. Thus, understanding of what a gaze behavior essential constitutes provides guidance to pervasive eye-tracking design.

Interaction essentials on the other hand are related to knowledge of user demands and expectations from the application. It is fundamental to understand what user expectations from the pervasive application are, and what the typical interaction patterns one can model in a pervasive application. Such knowledge guides requirements of cloud solution and applications, in the matter of speed, transmitted amount of data and response times.

Let's look at an example. Interactive uses of eye-tracking nowadays typically rely on large amounts of gaze data. For effective detection of interaction intent (Bednarik, Vrzakova, & Hradis, 2012), a database with training data need to be obtained and updated during the use to further improve the trained models and adapt them to new instances. Such analysis and processing is feasible on powerful machines and in the controlled environment of a laboratory. Out of the lab, as in case of our classroom example, these conditions and constrains will change, and so do essentials needed for functional interaction.

Instead of straightforward implementation based on current non-pervasive interaction standards, we call for deeper understanding of gaze and interaction essentials. We believe such knowledge will benefit the whole concept of pervasive eye-tracking along with technical improvements, like increased battery life or more compact hard drive.

Local vs. Distributed

Along the eye-tracking pipeline, starting with capture of the eye-image or with other way of eye-movement estimation, through fixation identification to processing of fixations and building user models, *computing* and *storage demands* differ. An important part of the future agenda will be to carefully evaluate the possibilities of the technology to deliver fast and accurate predictions and to handle the voluminous data. While raw data capture cannot be offloaded to the cloud, other parts of the pipeline can.

Especially in the context of mobile device interaction, some attempts already exist to integrate eye-tracking and cloud technologies. (Kao, Yang, Fan, Hwang, & Huang, 2011) present a modular view on eye-tracking pipeline and optimise the available resources of each mobile client. Because individual mobile devices do not have sufficient computing power, cloud computing can mitigate this challenge. In their proposal, a distributed computation carries on feature tracking and gaze mapping, while the initial stages are handled locally.

We expect similar considerations will become a common part of future pervasive eye-tracking interaction. In those contexts, one will need to decide how fast a query to the user model can be replied because gaze essential is carried out by segments of varying lengths. For example, in our current research we deal with short tree-fixations segments (Bednarik, Vrzakova, & Hradis, 2012) or with rather long fifteen-seconds segments (Bednarik, Eivazi, & Hradis, 2012). The different response times will have an implication of whether the data have to be processed locally or in cloud.

Pervasive Gaze Interaction in Cloud

Expectations from seamless pervasive eye-tracking interaction will be high, and pervasive cloud solutions will force designers and developers at every level to deepen their understanding of visual attention role and its relation to the application purpose. The future interactive applications need to guarantee:

- Free viewing without Midas touch and interaction slips (Vrzakova & Bednarik, 2013)
- Intelligent information display management
- Immediate interaction response
- Wire-free hardware solution
- Seamless and ready-to-go (re)calibration

The expectations drive requirements on all levels, from those related to design concepts, through the hardware improvements to application development.

Scientific progress supported by cloud-based eye-tracking

Cloud-based solutions for eye-tracking are not only enablers of technological advances. Several shortcomings of the state-of-the-art of eye-tracking research will be solvable when the distributed technologies become operational. The past eye-tracking studies were conducted in settings that do not correspond to future uses as imagined in our example. In other domains, such as language development, the necessity to collect data in realistic settings have long been recognized (Bruner, 1985). Eye-tracking studies, at the moment, still capture data in short experiments, in serialized manner where a limited number of participants are recorded separately, and in laboratory or otherwise constant settings. Development of the theory and methodology is thus confined to the settings in which the data is collected, and often eye-tracking studies resort to fishing trips to obtain insights (Holmqvist et al., 2011). The possibility for distributed gaze-data collection will allow eye-tracking research to free itself from the laboratory.

Second, while natural stimuli studies are abundant in eye-tracking literature, no understanding is available about individuals' visual attention in natural environments from a *long-term perspective*. The effect of task on eye- movement is profound (Hayhoe & Ballard, 2005; Yarbus & Riggs, 1967), yet, the long-term gap contains questions such how attention is situated, how it is particular or general with regard in in time. This is a crucial information for future ubiquitous systems to intelligently react on the users' behavior and tap into users' needs. To be able to adapt to the needs, changing physical and psychological conditions, and environments, the visual attention situatedness needs to be isolated, characterized and computationally modeled.

Third, no knowledge is available on visual attention of very large groups of users engaged in a particular interaction context. Finding whether people engage in a variety of behaviors in a given context or they converge in their behaviors is a common theme of eye-tracking research, as exampled by studies of expertise (Memmert, 2009; Gegenfurtner, Lehtinen, & Säljö, 2011; Bednarik, 2012). More research is however necessary to find the ranges, the nuances, and the cores of behavior in typical interactive contexts to provide trustworthy parameters of the models for best individual adaptation based on a large corpora. The performance of intelligent systems for individuals will be improved based on the massive visual attention datasets of numerous other users.

Conclusion and Future Work

The opportunity for cloud-based gaze applications is arriving. With no feasible solution at hand, the eye-tracking community needs to consider the challenges, requirements but also new pathways connected with the opportunity. The new domain of EyeCloud is complex; to be able to deliver really pervasive solutions our community needs to focus on a large gamut of issues ranging from HCI aspects, through parallel computing, embedded and network systems and hardware optimization, to privacy and ethical issues.

We chartered some of the borders and roads on the developing map of pervasive interactive eye-tracking. Specifically we pointed out those aspects that are relevant to implicit uses of gaze. Many other important topics, such as ethical and privacy issues, along with other social implications of pervasive cloud-based eye-tracking, need to be carefully examined in future research.

Acknowledgement

We would like to thank our colleagues at University of Eastern Finland, especially to Shawn Xiaolei, Chung Dao, Shahram Eivazi and Tersia Gowases for discussions leading to formulation of the EyeCloud concept and we appreciate the helpful comments from the PET-MEI reviewers.

References

Bednarik, R. (2012). Expertise-dependent visual attention strategies develop over time during debugging with multiple code representations. *International Journal of Human-Computer Studies*, *70*(2), 143–155.

Bednarik, R., Eivazi, S., & Hradis, M. (2012). Gaze and conversational engagement in multiparty video conversation: an annotation scheme and classification of high and low levels of engagement. In *Proceedings of the 4th workshop on eye gaze in intelligent human machine interaction* (p. 10).

Bednarik, R., Vrzakova, H., & Hradis, M. (2012). What do you want to do next: a novel approach for intent prediction in gaze-based interaction. In *Proceedings of the symposium on eye tracking research and applications* (pp. 83–90).

Bruner, J. (1985). Child's talk: Learning to use language. *Child Language Teaching and Therapy*, *1*(1), 111–114.

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: a meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, *23*(4), 523–552.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in cognitive sciences*, *9*(4), 188–194.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Weijer, J. Van de. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Kao, C.-W., Yang, C.-W., Fan, K.-C., Hwang, B.-J., & Huang, C.-P. (2011). An adaptive eye gaze tracker system in the integrated cloud computing and mobile device. In *Machine learning and cybernetics (icmlc), 2011 international conference on* (Vol. 1, pp. 367–371).

Lee Green. (2012). *Smart board tutorial.* `http://bit.ly/XPdLNc`.

Memmert, D. (2009). Pay attention! a review of visual attentional expertise in sport. *International Review of Sport and Exercise Psychology*, *2*(2), 119–138.

Van Krevelen, D., & Poelman, R. (2010). A survey of augmented reality technologies, applications and limitations. *International Journal of Virtual Reality*, *9*(2), 1.

Vrzakova, H., & Bednarik, R. (2013). That's not norma(n/l): a detailed analysis of midas touch in gaze-based problem-solving. In *Chi '13 extended abstracts on human factors in computing systems* (pp. 85–90). New York, NY, USA: ACM.

Yarbus, A. L., & Riggs, L. A. (1967). *Eye movements and vision* (Vol. 2) (No. 5.10). Plenum press New York.