Neurocomputing 207 (2016) 653–668

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Predicting task from eye movements: On the importance of spatial distribution, dynamics, and image features

Jonathan F.G. Boisvert, Neil D.B. Bruce *

Department of Computer Science, University of Manitoba, 66 Chancellors Circle, Winnipeg, Manitoba, Canada

ARTICLE INFO

Article history:
Received 10 November 2015
Received in revised form
13 May 2016
Accepted 15 May 2016
Communicated by Hantao Liu
Available online 25 May 2016

Keywords:
Eye movements
Fixations
Task prediction
Features
Visual search
Yarbus

ABSTRACT

Yarbus' pioneering work in eye tracking has been influential to methodology and in demonstrating the apparent importance of task in eliciting different fixation patterns. There has been renewed interest in Yarbus' assertions on the importance of task in recent years, driven in part by a greater capability to apply quantitative methods to fixation data analysis. A number of recent research efforts have examined the extent to which an observer's task may be predicted from recorded fixation data. This body of recent work has raised a number of interesting questions, with some investigations calling for closer examination of the validity of Yarbus' claims, and subsequent efforts revealing some of the nuances involved in carrying out this type of analysis including both methodological, and data related considerations. In this paper, we present an overview of prior efforts in task prediction, and assess different types of statistics drawn from fixation data, or images in their ability to predict task from gaze. We also examine the extent to which relatively general task definitions (free-viewing, object-search, saliency-viewing, explicit saliency) may be predicted by spatial positioning of fixations, features co-located with fixation points, fixation dynamics and scene structure. This is accomplished in considering the data of Koehler et al. (2014) [30] affording a larger scale, and qualitatively different corpus of data for task prediction relative to existing efforts. Based on this analysis, we demonstrate that both spatial position, as well as local features are of value in distinguishing general task categories. The methods proposed provide a general framework for highlighting features that distinguish behavioural differences observed across visual tasks, and we relate new task prediction results in this paper to the body of prior work in this domain. Finally, we also comment on the value of task prediction and classification models in general in understanding facets of gaze behaviour.

1. Introduction

Early seminal work in analyzing eye movement patterns by Buswell [9] and Yarbus [58] remains influential in shaping scientific discourse addressing the role of task in gaze behaviour. This includes the notion that an observer's task may be inferred from examining their eye movements.

There is a rich literature on research efforts targeted at predicting human gaze patterns, in most instances in the absence of a specific task Bruce et al. [7], Shen and Zhao [41], Han et al. [24], Loyola et al. [35], Borji et al. [3], Kümmerer et al. [31], Wilming et al. [55].

Contemporary research efforts have further examined the interaction between task and fixations, in some cases directly considering Yarbus' claims about the predictability of task from fixations [10,48]. Some of this analysis has leveraged modern techniques in pattern classification to directly predict task from recorded fixation data. One prior effort modelled heavily on the methodology of Yarbus' experiments, [19], considers three different classifiers applied to aggregate eye-movement statistics for task prediction. None of these classifiers yielded performance above chance in considering the aggregate fixation features. More recent work considering the same data, but instead using low-resolution fixation density patterns [1], or other statistics achieved above chance performance by revising the feature set and inference methods used. Further improvements for the same data set have also been achieved in assuming knowledge of the participant viewing the image, or the specific image under consideration [29]. Accounting for covert attention in models of this variety has also lead to higher prediction accuracies [22]. The data provided by

* Correspondence to: E2-408 EITC, Department of Computer Science, University of Manitoba, Winnipeg, MB, Canada R3T 2N2.
E-mail addresses: umboisve@myumanitoba.ca (J.F.G. Boisvert), bruce@cs.umanitoba.ca (N.D.B. Bruce).

http://dx.doi.org/10.1016/j.neucom.2016.05.047
0925-2312/

Greene et al. [19] has also inspired analysis of the possible impact of tasks requiring both visual processing and language processing [13]. There are evidently important individual differences that factor into viewing patterns [44], and such results also point to the notion that features such as fixation position and duration carry significant information about a viewers task.

In the context of attention modeling, there are many studies that address the relative contribution of *bottom-up* vs. *top-down* influences on the deployment of overt attention and fixation patterns. In the context of this paper, bottom-up refers to exogenous attention that is driven by properties of the visual stimulus, and independent of task or semantics. In referring to top-down processes, we refer to endogenous aspects of attention and guidance of gaze that are under executive control and may involve the influence of task directives, and working memory.

The spirit of Yarbus' assertions are closely related to an active overt attention process, drawing heavy influence from top-down bias. The critical importance of task in attention, and overt attention specifically is well supported as discussed in a number of recent studies and review papers addressing the relative importance of task and top-down cues [25,12,40,27,47,57,46,2].

While recent efforts leave little doubt that observed fixations provide a window into cognitive state or task directives, there is remaining benefit in examining task predictability from gaze statistics. One evident benefit from an applied perspective is the capability to infer task or intentions from fixations for applications in human machine interaction and human centric computing. Additional benefits of a more general nature arise from examining the ease with which different tasks may be distinguished based on gaze patterns, and in determining which features successfully discriminate between tasks. The degree of task separability has value in understanding similarity among visual and attentive mechanisms recruited for different tasks. Determining specific factors that distinguish tasks also points to targets for more careful examination in targeted experimental studies. In the domain of task prediction, prior work has focused heavily on confirming or denying the hypothesis that task may be predicted from gaze. For this reason, there has been a strong emphasis on prediction accuracy with less consideration of the role of different image or gaze related statistics in determining prediction performance. The work presented in this paper expands on the body of research involving task prediction in addressing the following important questions:

1. *Do relatively coarse grained tasks present distinct gaze statistics*? While several different sets of tasks have been examined in the literature, we consider task definitions that reside at a relatively general or coarse-grained level of abstraction. This serves to contribute to the growing body of efforts examining task prediction, and also to add diversity in the types of task sets considered.
2. *What methodological considerations are most critical to drawing value from efforts in task prediction*? If the goal of task prediction is to achieve something beyond confirming or denying Yarbus' assertions, there is value in ensuring that methodology allows for analysis beyond comparing prediction accuracies. We therefore employ methods for task prediction that are amenable to considering the relative importance of associated gaze and image related statistics, and discuss additional considerations of importance at the level of methodological details.
3. *Which gaze statistics are most important*? In considering a set of relatively coarse grained task directives and choosing suitable methods, we aim to establish which gaze statistics or image derived features seem to diverge most across different task definitions. This provides insight into information represented within different types of gaze or image related statistics.
4. *What does task prediction tell us (and not tell us) about vision*? Task prediction establishes that different tasks may be distinguished on the basis of gaze statistics. Results presented in this paper further reveal the relative importance of different types of features for the coarse-grained tasks examined in this paper, and also within existing studies. However, it is important to address the limitations on what studies in task prediction are able to convey about human vision. An additional goal of this paper is therefore to establish what benefits and limitations exist in examining task specific behaviour within a task prediction paradigm.

These four questions form the core of motivation for the work presented in this paper, and its novelty. First, we consider a very different set of task defined gaze data than has been considered previously, providing some new observations about challenges in this problem domain. More importantly though, we also explore in detail methodological considerations for approaching work in predicting X from gaze, where X might correspond to an assigned task, affective response to images or environment, or any other measurable factors. This is achieved in highlighting the importance of considering different types of features, and associated subsets while also applying predictive methods that offer feedback on the relative value of said features.

The balance of the paper is structured as follows: in Section 2 we present a survey of studies that emphasize task prediction highlighting differences in the set of tasks considered, methods and accuracies achieved across these studies. Section 3 presents the experimental methods that are exercised in this paper. This includes further details on the dataset, and types of features considered for task categorization. Following this, in Section 4, we present the relative classification performance that is achieved in considering the spatial distribution of fixations, local features at fixated locations, fixation dynamics and global scene structure. This analysis considers different conditions, including pooled fixation data across all observers, as well as fixations for single observers subject to different methods of partitioning the image set. Various combinations of 4-way, 3-way and binary classifications are considered where appropriate to shed further light on factors that separate tasks. We discuss the broader implications of this analysis in Section 5, including limitations and possible fruitful directions forward. Finally, Section 6 summarizes important results from this paper in addressing the role of task in observed fixation behaviour.

2. Prior work in task prediction

There are a variety of recent studies that consider this problem with direct reference to Yarbus' work, or specifically involving a classification paradigm for assessing the predictability of task from fixation data. DeAngelus and Pelz [15] re-examined Yarbus' work, including tools, methods, and implications of Yarbus' findings. They also replicated Yarbus' original experiment using updated methods and a larger pool of participants and paintings for fixation recording. Their results demonstrated patterns consistent with Yarbus' data for Repin's painting using modern eye tracking devices, while restricting observations to a shorter time course. Castelhano et al. [10] showed that an observer's task (object search and memorization) influences eye movement behaviour at the level of fixation durations and saccade amplitudes, specifically at the level of aggregate eye movement measures rather than individual fixation or saccade statistics. In the case of the memorization task, a larger area of the image was fixated and while the average fixation duration did not vary significantly between tasks, certain areas were re-fixated (approximately) increasing their total