Vision Research 108 (2015) 67–76

Contents lists available at ScienceDirect

Vision Research

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

Humans have idiosyncratic and task-specific scanpaths for judging faces

Christopher Kanan [a,*], Dina N.F. Bseiso [b], Nicholas A. Ray [b], Janet H. Hsiao [c], Garrison W. Cottrell [b]

[a] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
[b] Department of Computer Science and Engineering, University of California, San Diego, La Jolla, CA, USA
[c] Department of Psychology, University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:
Received 14 May 2014
Received in revised form 3 October 2014
Available online 30 January 2015

Keywords:
Eye movements
Machine learning
Scanpath routines
Face perception

ABSTRACT

Since Yarbus's seminal work, vision scientists have argued that our eye movement patterns differ depending upon our task. This has recently motivated the creation of multi-fixation pattern analysis algorithms that try to infer a person's task (or mental state) from their eye movements alone. Here, we introduce new algorithms for multi-fixation pattern analysis, and we use them to argue that people have scanpath routines for judging faces. We tested our methods on the eye movements of subjects as they made six distinct judgments about faces. We found that our algorithms could detect whether a participant is trying to distinguish angriness, happiness, trustworthiness, tiredness, attractiveness, or age. However, our algorithms were more accurate at inferring a subject's task when only trained on data from that subject than when trained on data gathered from other subjects, and we were able to infer the identity of our subjects using the same algorithms. These results suggest that (1) individuals have scanpath routines for judging faces, and that (2) these are diagnostic of that subject, but that (3) at least for the tasks we used, subjects do not converge on the same "ideal" scanpath pattern. Whether universal scanpath patterns exist for a task, we suggest, depends on the task's constraints and the level of expertise of the subject.

1. Introduction

Multi-fixation pattern analysis (MFPA) is a new eye movement analysis technique that harnesses machine learning to make inferences about subjects from their eye movements (Benson et al., 2012; Greene, Liu, & Wolfe, 2012; Tseng et al., 2013; Kanan et al., 2014; Borji & Itti, 2014). MFPA algorithms take a person's scanpath, a sequence of fixations, as their input and use the scanpath to infer traits such as the task the person was given. If an algorithm can make this inference above chance when trained on a person's scanpaths for specific tasks, then this suggests that the person might have scanpath routines for accomplishing one or more of the tasks. Prior work with MFPA has focused on validating the technique for inferring the task given to a subject when viewing scenes (Greene, Liu, & Wolfe, 2012; Kanan et al., 2014; Borji & Itti, 2014) and for inferring whether the subject has a particular disease (Benson et al., 2012; Tseng et al., 2013). In this paper, we use MFPA to determine if people have scanpath routines for making different inferences about faces.

Humans make about three saccadic eye movements per second. Saccades are needed because the human retina has variable spatial resolution. It only acquires high resolution information in its central (foveal) region, with the resolution in the retinal periphery being far lower. The information in the periphery, along with information about the task being performed, can help direct saccades to diagnostic features for the task at hand. It makes sense, then, for humans to deploy scanpath routines for specific tasks so that diagnostic information can be acquired using as few fixations as possible. Formally, we define a scanpath routine as a task-specific sequence of fixations that exhibits a particular repeated spatial or spatio-temporal pattern. In order to rule out certain trivial cases, we also require that scanpath routines be acquired implicitly through learning, rather than elicited via direct instruction.

We hypothesized that scanpath routines for making common inferences about faces are likely to exist because some regions of the face are more diagnostic than others for some tasks. For example, the mouth is more diagnostic when judging whether a face is expressive or not and the eyes are crucial features for judging gender and identity (Gosselin & Schyns, 2001; Schyns, Bonnar, & Gosselin, 2002). Similarly, in face recognition it has been shown that the left eye is the most diagnostic feature initially, followed by both eyes (Vinette, Gosselin, & Schyns, 2004). This finding is corroborated by results showing that people tend to fixate slightly

* Corresponding author.
E-mail addresses: ckanan@caltech.edu (C. Kanan), dbseiso@ucsd.edu (D.N.F. Bseiso), niray@ucsd.edu (N.A. Ray), jhsiao@hku.hk (J.H. Hsiao), gary@ucsd.edu (G.W. Cottrell).

http://dx.doi.org/10.1016/j.visres.2015.01.013
0042-6989/

to the left of the nose initially during face recognition (Hsiao & Cottrell, 2008). Taken together, these findings indicate that different face regions have varied diagnostic utility. However, the scanpaths people use to make inferences from faces may not be universal, because people have different experiences and slightly different visual systems. Peterson and Eckstein (2012) showed that the fixation points used by people to determine age, gender, and emotional state of a face differ across these three tasks. In subsequent work, they also showed that human eye movements during face identification were idiosyncratic (Peterson & Eckstein, 2013). Mehoudar et al. (2014) similarly found that scanpaths during face viewing are idiosyncratic, and that individuals continued to use the same idiosyncratic patterns when viewing faces 18 months later. Finally, several papers have shown that people's scanpaths have different properties when viewing novel faces compared to viewing familiar ones (Althoff & Cohen, 1999; Joyce, 2000).

The idea of scanpath routines is closely related to "scanpath theory" (Noton & Stark, 1971; Spitz, Stark, & Noton, 1971). Scanpath theory argues that eye movements are generated in a top-down manner to facilitate correct recognition of an image by comparing it to previous experience. Learning a recognition task is taken to mean storing both the visual features and the motor sequence used to acquire the features. Recognition involves recapitulating the same scanpath when encountering the same stimulus. The strong form of scanpath theory predicts that individuals should deploy an identical pattern of eye movements during recognition, which is not consistent with human behavior (Henderson, 2003). If humans did behave according to the strong theory, then it is likely that this behavior would have limited utility since humans rarely encounter exactly the same visual stimulus twice. A more general version of scanpath theory would predict that eye movements should be similar (statistically regular) between viewings of images from the same stimulus class, and this theory would allow scanpath routines to evolve over time to enable improved processing of the stimulus class, e.g., doing the task accurately using fewer fixations. This version of scanpath theory is consistent with our notion of scanpath routines.

To demonstrate the existence of scanpath routines for specific tasks, it is necessary to show that scanpaths are altered by the task, but this is not sufficient because it allows certain trivial cases. For example, Tatler et al. (2010) showed participants a photo of Alfred Yarbus wearing a coat and asked subjects various questions. When the subjects engaged in free viewing, the majority of their fixations were located on Yarbus's face, whereas when they were instructed to remember his clothing their fixations were more evenly distributed between his face and clothes. In their study, the instructions essentially required the subjects to view different parts of the image (the clothes). Obviously, one can easily create a situation where verbal instructions result in discriminable scanpaths. Trivially, one can ask the subject to look at the upper left hand corner of the image on one trial and the lower right on another. These examples are not scanpath routines because they are not acquired implicitly, i.e., the instruction tells the subject where to look.

For an observer to deploy a scanpath routine for a task, we believe two constraints must be met. First, the task needs to be one that an observer has experience with. Second, the task should be one in which the same task-diagnostic regions in each image will need to be fixated to perform the task accurately and using as few fixations as possible. In our study, these conditions are satisfied by using aligned facial images, such that the information is always in relatively the same locations in each image, and asking questions about them that subjects are likely to have experience with. From birth, humans acquire an enormous amount of experience in making judgments about others from their faces, suggesting that they will have established scanpath routines for efficiently answering these questions about faces. We ask our subjects to judge the age, fatigue, angriness, happiness, trustworthiness, and attractiveness of the people in the images. Because informative facial features are always located in the same relative position, it is possible to develop a scanpath routine so that task-relevant information can be obtained. To adequately test for scanpath routines, we ask subjects the same six questions about every image, so that we are able to determine that the eye movements are not driven purely by the stimulus.

We attempt to find evidence for scanpath routines when making judgments about faces by using three different MFPA algorithms. The first method uses only summary statistics, i.e., the mean number of fixations in a trial and mean fixation duration, to make its inference. This approach ignores the spatio-temporal dynamics of the fixations, but it serves as a useful baseline. The second algorithm models the spatial distribution of fixations, including their duration at each location, but it ignores the temporal order information. The last algorithm is able to preserve information about the order of fixations as well as use spatial information. If any of the methods is above chance then we have strong evidence for scanpath routines for judging faces. By comparing the spatial and spatio-temporal methods, we can gain insight into the nature of these scanpath routines. For instance, if the spatio-temporal algorithm is significantly more accurate than the spatial algorithm then we can infer that there are diagnostic temporal regularities in the scanpaths.

We look for evidence of scanpath routines using a within-subjects and between-subjects analysis. If our algorithms are less accurate in a between-subjects analysis compared to within-subjects, then this suggests that people have idiosyncratic scanpath routines and that we can use our algorithms to infer the identity of the subjects.

2. Methods and materials

2.1. Participants

The data used in our experiments is from 12 male and 12 female Caucasian UCSD students (mean age 19 years 8 months; age range 18–22), who received course credit for their participation. One additional female was recruited, but the data was excluded because she did not respond before the timeout in 98% of the trials. For the other participants, this occurred in 3% of trials on average (Min: 0%, Max: 11%). This data was not excluded in our analysis. All participants were right-handed based on the Edinburgh handedness inventory (Oldfield, 1971), and all had normal or corrected-to-normal vision. Participants gave informed consent after the study had been explained to them, the study was approved by UC San Diego's Institutional Review Board, and the work was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

2.2. Stimuli

The stimuli in our experiment consisted of 48, 561×701 pixel color face images, half female and half male. There were three images of each face model in this dataset, expressing either happy, angry, or neutral facial expressions. During a brief familiarization session to acquaint the participants with the experimental paradigm (described further in Section 2.4), five additional images of face models not used in the remainder of the study were employed. All face images were front-view, Caucasian, and none had facial hair or glasses. The images came from four age groups: child, young adult, adult, and elderly. Because no single dataset at the time of the study contained both images of children and elderly individuals, we combined images from two face datasets. Images