Spatial Vision, Vol. 22, No. 2, pp. 161–177 (2009)

Also available online - www.brill.nl/sv

DOVES: a database of visual eye movements

IAN VAN DER LINDE [1,2,*], UMESH RAJASHEKAR [1,3], ALAN C. BOVIK [1] and LAWRENCE K. CORMACK [1]

[1] *Center for Perceptual Systems, Department of Psychology, University of Texas at Austin, Austin, TX 78712, USA*
[2] *Department of Computing, Anglia Ruskin University, Bishops Hall Lane, Chelmsford CM1 1SQ, UK*
[3] *Laboratory for Computational Vision, New York University, New York, NY 10003, USA*

Received 24 August 2007; accepted 24 January 2008

Abstract—DOVES, a database of visual eye movements, is a set of eye movements collected from 29 human observers as they viewed 101 natural calibrated images. Recorded using a high-precision dual-Purkinje eye tracker, the database consists of around 30 000 fixation points, and is believed to be the first large-scale database of eye movements to be made available to the vision research community. The database, along with MATLAB functions for its use, may be downloaded freely from http://live.ece.utexas.edu/research/doves, and used without restriction for educational and research purposes, providing that this paper is cited in any published work. This paper documents the acquisition procedure, summarises common eye movement statistics, and highlights numerous research topics for which DOVES may be used.

Keywords: Eye tracking; eye movements; database; natural image statistics; point-of-gaze statistics.

INTRODUCTION

Since the human visual system evolved in a natural environment, and because natural images occupy a relatively small subspace of all possible images, it is theorized that early visual processing exploits the statistical biases inherent in our visual surroundings (Barlow, 1961; Párraga *et al.*, 2000; Simoncelli and Olshausen, 2001). A body of work has focussed on exploring the relationship between the statistics of natural scenes and the structure of neural computations (Field, 1987; Hancock *et al.*, 1992). To ensure the accurate measurement of natural image statistics, great care has been taken to create and use calibrated natural image datasets; these have been used successfully to model the behaviour of complex cells

*To whom correspondence should be addressed. E-mail: ianvdl@ece.utexas.edu

in V1 by imposing statistical constraints on the processing of local image patches (van Hateren and van der Schaaf, 1998).

It is well known that eye movements are also an integral part of the encoding of visual stimuli; human eyes actively interact with their visual environment, gathering information from the *foveated* (variable spatial resolution) visual input using a combination of steady eye fixations linked by rapid ballistic eye movements called saccades (Henderson, 2003, 2007; Rayner, 1998; Yarbus, 1967). In addition to analyzing the relationships between natural scene statistics and the structure of neural computations, an understanding of how observers select and sequence image regions for foveal scrutiny is necessary to gain a complete understanding of the human visual system (Findlay and Gilchrist, 2003).

Recent studies have greatly improved our understanding of how eye movements are deployed in real-world scene viewing (Henderson, 2003; Torralba *et al.*, 2006), during reading and information processing (Rayner, 1998), visual search (Najemnik and Geisler, 2005), and during natural tasks requiring coordinated eye and body movements (Hayhoe and Ballard, 2005). Indeed, the study of eye movements is a burgeoning research area with a spectrum of applications ranging from cognitive psychology to computation, neuroscience and business. Despite this, fundamental questions relating to how image loci are selected for fixation remain unanswered: To what degree are eye movements inherently biased? To what degree are eye movements affected by different image properties?

In the field of gaze modelling, there is increasing interest in computing natural scene statistics directly at the *point of gaze* of observers, and thereafter establishing the degree to which the statistical properties of image features at observers' fixations differ from regions selected at random. The availability of relatively inexpensive eye trackers has made this approach feasible. In one such study Reinagel and Zador (1999) showed that the regions around human fixations have higher spatial contrast and spatial entropy compared to random regions, indicating that human eye movements may be deployed to select image regions that help maximise the information transmitted to the visual cortex by minimising the redundancy in the image representation. Similar findings for other image statistics have been reported by several other researchers (Einhäuser and König, 2003; Itti and Koch, 2000; Parkurst and Niebur, 2003, 2004; Parkurst *et al.*, 2002; Privitera and Stark, 2000; Rajashekar *et al.*, 2007; Tatler *et al.*, 2006), complemented by alternative approaches focussing on top-down/contextual fixation guidance mechanisms (Torralba, 2003; Torralba *et al.*, 2006). Models of top-down and bottom-up mechanisms are not necessarily in conflict, with several researchers proposing that each may have precedence at different times, i.e. bottom-up effects are more important soon after stimulus onset where top-down knowledge is unavailable (Li and Snowden, 2006), and during particular visual tasks, e.g. more so in recognition memory rather than search (Underwood and Foulsham, 2006; Underwood *et al.*, 2006). Further, it has been proposed that information from both bottom-up and top-down mechanisms may be ultimately united in the