

DOVES: a database of visual eye movements

IAN VAN DER LINDE^{1,2,*}, UMESH RAJASHEKAR^{1,3}, ALAN C. BOVIK¹
and LAWRENCE K. CORMACK¹

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

visual stream (Treue, 2003; vanRullen, 2005), possibly using a *saliency map* that weights the degree of attraction at each location in the visual field relative to its neighbours (Li, 2001). The study of image features at point of gaze enables the creation of *artificial saliency maps*: matrices that express the computed likelihood of each region being selected for fixation by new observers (Itti and Koch, 2000). This enables the prediction of fixations on novel scenes, and may be evaluated *post-hoc* by comparison with the actual fixation loci of human observers. Such experiments are typically conducted with either a small number of observers, and/or a small collection of uncalibrated images. To the best of our knowledge, to date, there exists no publicly available database of high precision eye movement recordings using natural calibrated images involving a large number of human observers.

In this paper, we present DOVES: a database of visual eye movements, which provides recordings of eye movements for 29 observers as they viewed 101 calibrated natural images (van Hateren and van der Schaaf, 1998) using a high-accuracy dual-Purkinje eye tracker. The database, which consists of around 30 000 fixation points, was collected using a large group of human observers, a large, carefully selected set of high resolution natural calibrated images, and was recorded at very high spatial and temporal resolution. In the Methods section, we provide details of the experimental apparatus, image database, data collection procedure and visual task.

We believe that the availability of a large eye movement database will be useful to the vision research community in a number of contexts. First, such a database provides a benchmark against which competing gaze modelling algorithms can be tested: the fixation loci in DOVES are both numerous, set upon calibrated images, and known to a high degree of accuracy. This enables quantitative comparison with the salient regions identified by either established or fledgling gaze-modelling algorithms. Second, the fixations in DOVES may be used directly as input to novel saliency *learning* algorithms (Kienzle *et al.*, 2006). Third, DOVES presents a source of generic eye movement data for calculating probability distributions (such as those relating to fixation centrality (Tatler, 2007), fixation duration, saccade angle and saccade length) and sequential eye movement characteristics, enabling common *image-independent* eye movement statistics in a typical fixed-head display configuration to be established. Fourth, the database provides a source of genuine human-fixation patches that may be used as stimuli in studies of visual memory, or to stimulate visual cells in electrophysiology. The use of DOVES in these and other common vision scenarios is described in *Applications*, which also summarises the results of some initial studies.

METHODS

Observers

A total of 29 unpaid adult human volunteers (18 male and 11 female, mean age 27) served as observers. All observers either had normal or corrected-to-normal vision.

Observers consisted of members of the public, undergraduates, graduates students, research fellows and faculty from the University of Texas at Austin from a range of academic disciplines. A recent study documented that the eye movements elicited by observers during tasks such as reading were affected by cultural group (Rayner *et al.*, 2007). To enable data to be partitioned to study such group effects, participant details for DOVES available on-line include age, handedness and gender.

Each observer visited for a single session, and only two had seen the image stimuli previously; 24 were naïve as to the purpose of the experiment. All participants were briefed on the nature of the visual task they were to perform, and consented to participation in writing, in accordance with the applicable laws/regulations. The experiment was conducted observing University regulations for experiments involving human subjects.

Natural image stimuli

101 static images of size 1024×768 pixels (cropped from the centre of the original 1536×1024 images) were manually selected from a calibrated greyscale natural image database (van Hateren and van der Schaaf, 1998). Since the initial goal of our study was to evaluate the influence of low-level image features in drawing observers' fixations, images containing man-made structures and features such as animals, faces, and other items of particularly obvious semantic interest were omitted. Images whose luminance histograms suggested saturation of the capture device, and thus exhibited non-linearity, were also omitted. Typical images are shown in Fig. 1.

Stimuli were displayed on an Image Systems 21" greyscale gamma corrected monitor (Image Systems Corp. Minnetonka, MN), positioned at a distance of 134 cm from the observer, controlled by a Matrox Parahelia graphics card (Matrox Graphics Inc., Dorval, Quebec, Canada). The screen resolution was set at 1024×768 pixels, corresponding to about 1 arc min per pixel (60 pixels per degree of visual angle). The total spatial extent of the display was thus about $17^\circ \times 13^\circ$ visual angle (notwithstanding that the display window is slightly less than the CRT size, with minimal screen convexity). The MATLAB psychophysics toolbox (Brainard, 1997; Pelli, 1997) was used for stimulus presentation. Since the range of brightness varied drastically across the image database, each image was luminance adjusted such that the brightest point in each image corresponded to the brightest output level of the monitor, without affecting image contrast (i.e. distinct luminance levels were never pooled).

Before displaying each natural image stimulus, a full-screen Gaussian noise image was displayed to help eliminate after-images originating from the previous image that may have otherwise attracted fixations. The ambient illumination in the experiment room was kept low and constant for all observers, with a minimum of 5 min luminance adaptation provided while the eye-tracker was calibrated and practice session completed (see below).

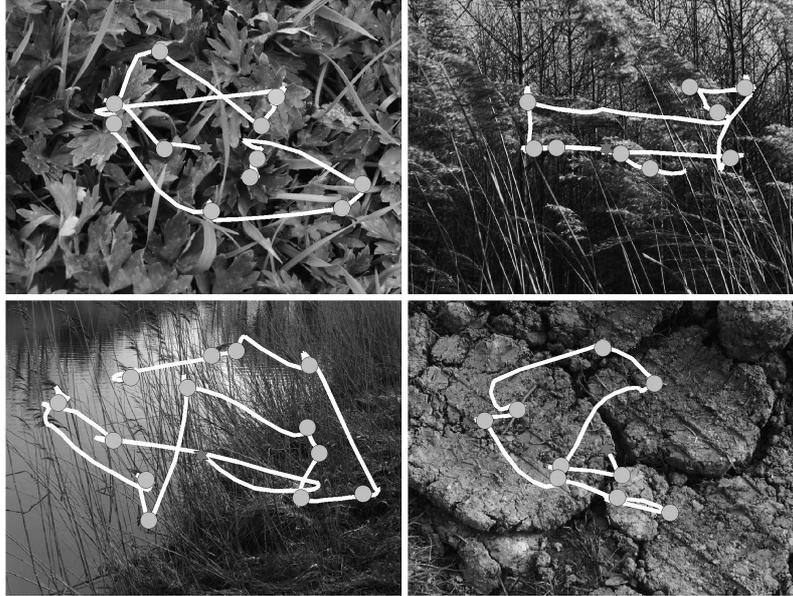


Figure 1. Examples of images used in the experiment. Also shown are an observer's eye movement trace superimposed on the image stimulus. The dots are the computed fixations. The star marker at the centre of the image is the first (forced) fixation.

Visual task

Observers were instructed to free-view each of the 101 images. Each was displayed for 5 s. To discourage observers from fixating at only one location, and to ensure a somewhat similar cognitive state across observers, a simple memory task was used: following the display of each image, observers were shown a small image patch (about $1^\circ \times 1^\circ$) and asked to indicate (using a handheld keypad) whether the image patch was from the image they just viewed (yes/no). Auditory feedback (*via* a sampled voice) was provided to indicate a correct or incorrect response. The probability of the patch having been part of the stimulus image was 50%. In the case that the patch was part of the stimulus image (a 'target'), it was sampled, unbeknownst to observers, at one of their preceding fixation coordinates (see Note 1). In the case that the patch was not from the stimulus image (a 'lure'), it was selected from an alternative set of 40 images of similar content. Since patches had a per-trial randomized 50% probability of being a target (really part of the image) or lure (originating from a different image), the possibility of learned probability matching was removed. To preclude the possibility of luminance-dominated matching, the brightness of the patch was jittered randomly. In order to preserve structure, the jitter magnitude was randomly selected from a range limited such that distinct luminance levels were never pooled. Observers were made aware of the brightness jittering during the training session (see below).

Before the main experiment, a training session of 10 trials was completed in order that observers could gain familiarity with the handheld response keypad, dark adapt, and become comfortable in the experimental environment prior to data collection. Images for the practice session were different from those used for the main experiment, but were selected from the same database.

The average duration for the experiment was approximately 1 h, including eye tracker calibration. Observers who became uncomfortable during the experiment were permitted to rest for any duration they desired. Post-experimental debriefing revealed that most observers rated the eye-tracker as only mildly uncomfortable. Plotting the mean performance of the observers over time for the visual task does not suggest a prevailing fatigue factor, with the task performance slope remaining constant throughout (Fig. 2), indicating, on average, a sustained level of concentration/effort. Behavioural results for the visual task are provided in *Memory-related experiments*.

Eye tracking

As the observers viewed each scene, their eye movements were recorded using a Fourward Technologies Generation V dual-Purkinje eye tracker (Buena Vista, VA). It has an accuracy of <10 min of arc, precision of about 1 min of arc,

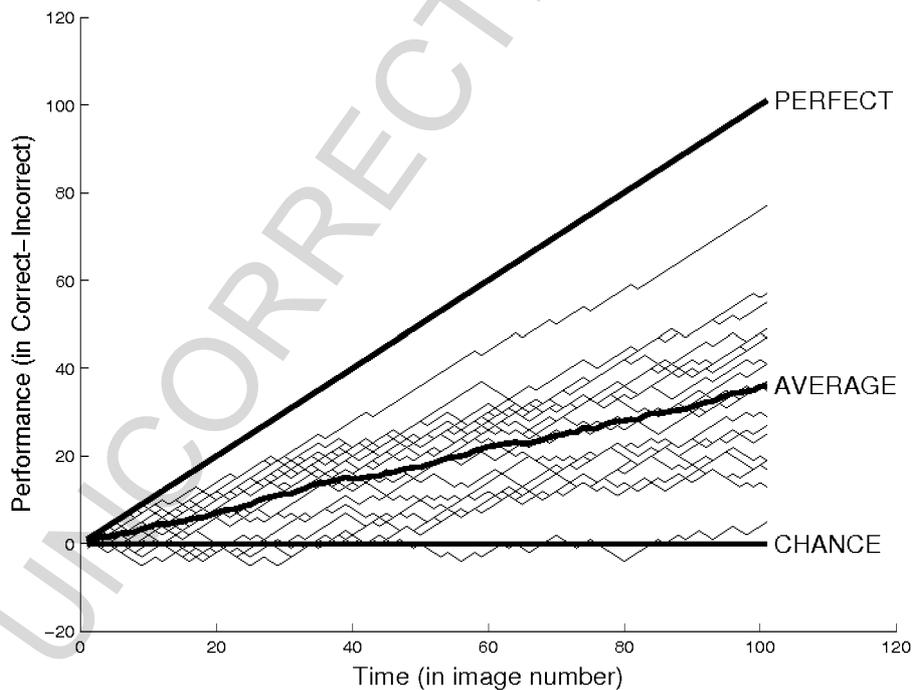


Figure 2. Subject performance as a function of number of images viewed. Performance was measured as the number of correct responses minus the number of incorrect responses to the visual task.

a response time of <1 ms, and bandwidth of DC to >400 Hz. The output of the eye tracker (horizontal and vertical eye position signals) was sampled at 200 Hz by a National Instruments data acquisition card (National Instruments Corporation, Austin, TX) in a Pentium IV host computer, where the data were stored for offline analysis.

Monocular eye tracking was used. A bite bar and forehead rest were used to restrict the subject's head movement. The subject was first positioned in the eye tracker and a system lock onto the subject's eye established. A linear interpolation routine using a 3×3 calibration grid was then completed to establish a transformation between the output voltages of the eye tracker and the position of the subject's gaze on the computer display. Calibration also accounted for crosstalk between the horizontal and vertical voltage measurements.

The above calibration routine was compulsorily repeated every 10 images, and a calibration test run after every image. This test was achieved by requiring that observers fixate for 500 ms within a 5 s time limit on a central square marker (about $0.3^\circ \times 0.3^\circ$) prior to progressing to the next image in the stimulus collection. If the calibration had drifted, the observer would be unable to satisfy this test, and the full calibration procedure was repeated. The average number of calibrations per observer for the 101 images was 16.5 ($SD = 4.7$), i.e. typically about 7 trials were completed before the calibration test was failed. The requirement for a central fixation prior to displaying the next image ensured that all observers commenced viewing the image stimuli from the same location.

Fixation generation

Sampled voltages corresponding to the eye movements of the observers for each trial were converted to gaze coordinates (i.e. position of gaze on the image in pixels). Next, the path of the subject's gaze was divided into fixations and the intervening saccadic eye movements using spatio-temporal criteria derived from the known dynamic properties of human saccadic eye movements. Stated simply, a sequence of eye position recordings was considered to constitute a fixation if the recorded gaze coordinates remained within a stimulus diameter of 1° visual angle for at least 100 ms. The exact algorithm (adapted from ASL, 1998) accounts for drifts, blinks and micro-saccades. A mean of 12.1 ($SD = 3.0$) fixations per image (Fig. 3A) was observed. Patterns of fixations for four single trials for one observer are shown in Fig. 1. The serpentine lines show the eye movement trajectories linking serial fixations (marked by circles). The first (forced) fixation is denoted by a star near the centre of the image. Since there are several techniques to compute fixations from eye movement traces, the online database also provides users with the raw eye movement data sampled at 200 Hz allowing the application of other fixation detection algorithms.

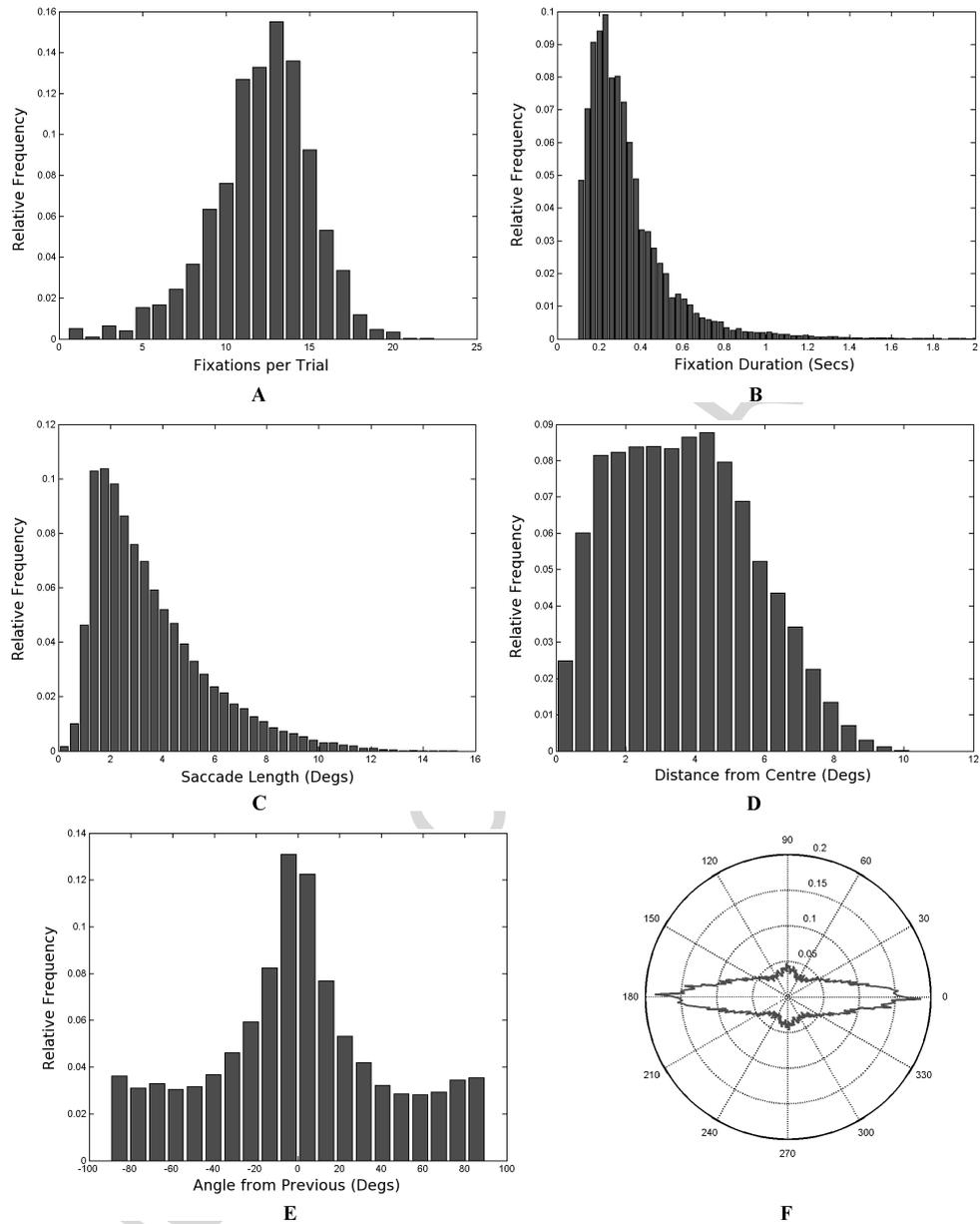


Figure 3. Histograms representing: (A) fixations per trial, (B) fixation duration, (C) saccade length, (D) distance from center, (E) saccade angle, (F) saccade angle (polar), (G) angle from centre and (H) angle from centre (polar).

APPLICATIONS

We envision that DOVES will be useful for several applications, a number of which are outlined below.

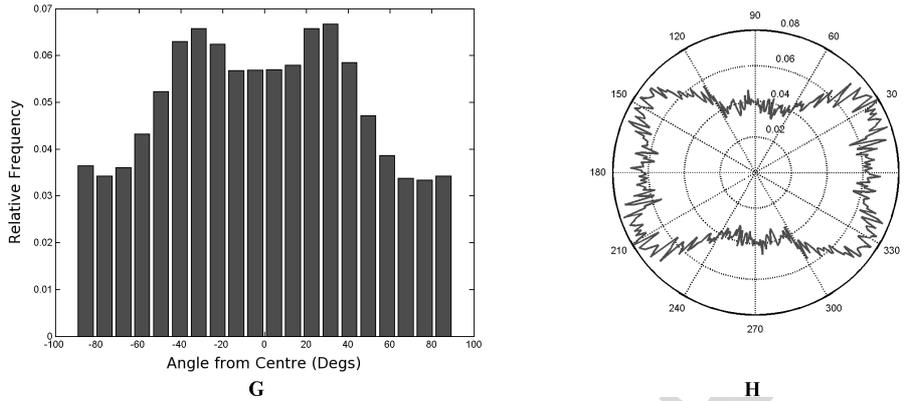


Figure 3. (Continued).

Gaze selection

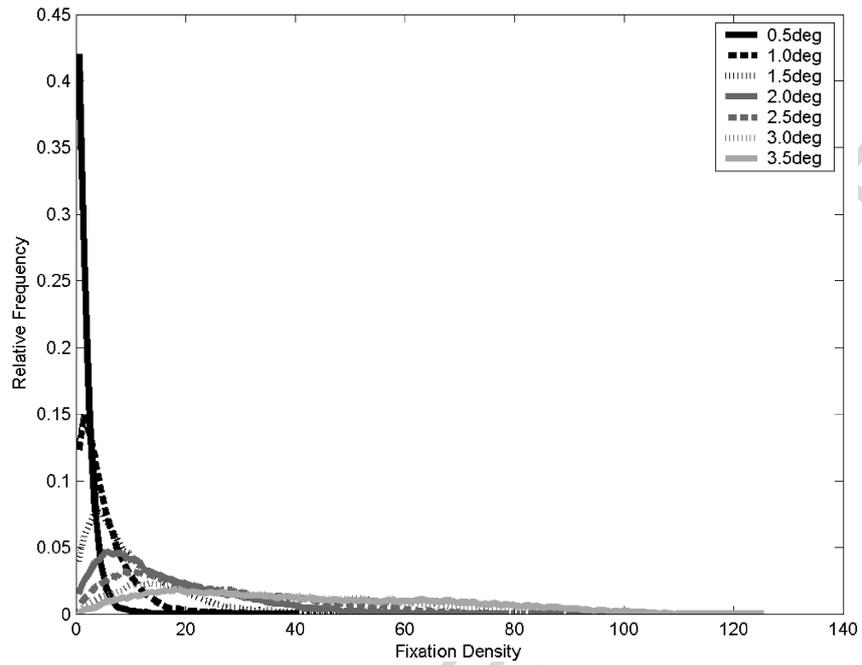
The ability to measure image statistics at an observer's point of gaze is useful in developing algorithms that select new fixation coordinates that mimic human gaze patterns. Using DOVES, we have studied the statistics of four low-level local image features: luminance, RMS contrast, and bandpass outputs of both luminance and contrast, and discovered that the image patches around human fixations had, on average, higher values of each of these features than the image patches selected at random. Using these features, we developed an algorithm to select fixations in novel scenes. The fixations selected by the algorithm were found to correlate well with observers' gaze positions (Rajashekar *et al.*, in press).

At present, it is difficult to compare existing models for gaze prediction due to lack of a common, large, accurate database of calibrated images with associated eye movements. By providing this information, DOVES can be used to evaluate the performance of competing algorithms by measuring the correlations between a model's prediction and the recorded fixations. DOVES also provides an opportunity to investigate the unique contributions of other potentially important low-level image features (Wolfe and Horowitz, 2004), for example, using multiple regression to identify uniquely contributing image statistics, or recent advances in machine-learning to identify image features that are elevated at patches centred at human fixations; for example, using the image patches at human fixations as input to self-learning algorithms that do not make any *a priori* assumptions about image structures/features that draw fixations (Kienzle *et al.*, 2006). Furthermore, the opportunity to analyze the correlation between eye movement statistics (such as fixation duration, or saccade length/angle) and the statistics of underlying image features exists, an aspect of gaze modelling that has, to date, received little attention. Demonstrating the validity of this approach, Tatler *et al.* (2006) recently found that frequency statistics at fixated regions produced by long *vs.* short saccades differ, with longer saccades tending to random frequency content.

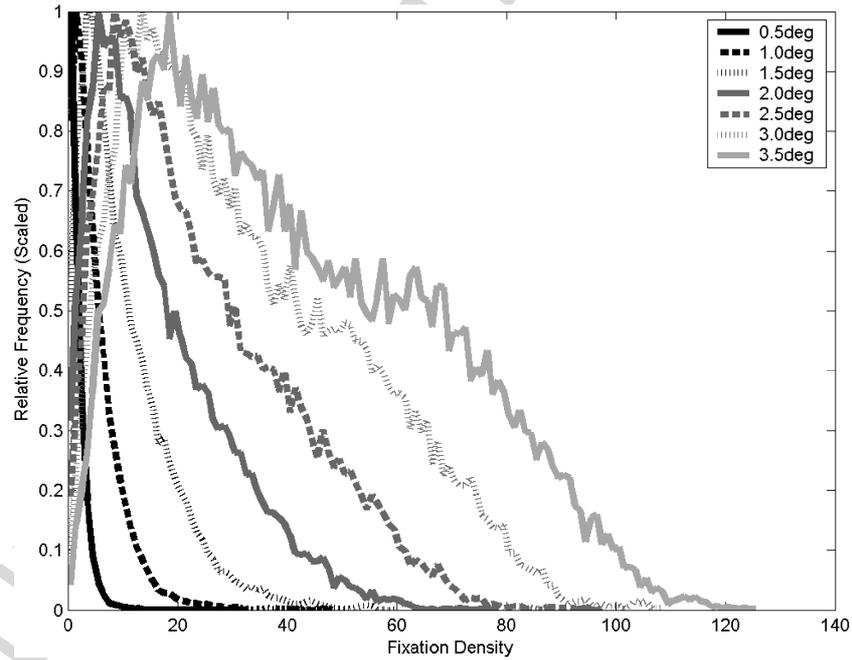
Eye movement patterns

Eye movement data may be subjected to a wide variety of statistics to determine saccade lengths, angles, speeds and fixation locations, durations and degrees of clustering, to name but a few. Whilst individual experiments may ascribe specific meanings to particular eye movement statistics, their general distributions are of great importance to applications such as accurate modelling/prediction of human gaze patterns, and in experimental procedures for which deviations from expected values in any of these measurements are to be interpreted as indicative of the observer's cognitive state. A small sample of statistics applied to the eye movement database are provided in Fig. 3: (A) number of fixations per 5 s trial, (B) fixation duration, (C) saccade length, (D) fixation distance from screen centre (where the screen centre represents both the forced fixation location at stimulus onset, and point of oculomotor equilibrium, i.e. centre forward gaze), (E) saccade angle (0° indicates a horizontal gaze shift with respect to the previous fixation, $-90/90^\circ$ indicate a vertical gaze shift), (F) saccade angle (polar), (G) fixation angle from screen centre (i.e. angle from 0 to 360° of each fixation with respect to the screen centre/fixation location at stimulus onset/point of oculomotor equilibrium) and (H) fixation angle from centre (polar). The high level of organisation in these eye movement measurements shows that sufficient fixations were recorded to infer statistical biases/trends with a degree of confidence. In gaze modelling/prediction, the degree to which the fixations of different observers are attracted to common image regions is also of great interest. Fixation density was calculated for a range of radii centred at the observer's fixations. For each fixation, the number of other fixations on the same image (from the same or other observers) within a square window was counted. The distribution of pooled fixation densities for ascending window sizes (from 0.5 to 3.5°) for all images is shown in Fig. 4, illustrating that as window size increases, higher fixation densities occur with considerably greater frequency. The first (forced) fixations on stimulus onset were omitted.

Clearly, eye movements exhibit several statistical biases that appear to be independent of the underlying stimulus (a selection of such statistics was presented above). For example, when presented with a novel scene, human observers tend to fixate first at the centre of the screen. It can also be shown that the magnitude of saccades follow a log-normal distribution (Fig. 3C), and that saccades are more likely to be deployed horizontally than vertically (Fig. 3E and 3F). DOVES can provide a useful framework for discovering and validating both known potential biases of human eye movements (such as the tendency of observers to fixate near the image centre/point of oculomotor equilibrium, Fig. 3D), and unknown biases (such as possible relationships between saccade magnitude and saccade angle). Finally, while there has been significant progress in modelling/predicting image locations likely to attract fixations, predicting the sequence of eye movements still remains a challenging task. The large collection of eye movement patterns in DOVES can be used to uncover common sequential patterns using such techniques as Kalman filters or Markov processes. For example, any tendency to follow large exploratory



(A)



(B)

Figure 4. Fixation densities for ascending radii (0.5 – 3.5°) centred at each fixation, sampled from respective source images, individually scaled by (A) sum and (B) max.

saccades by smaller investigatory saccades, decreases in inter-observer fixation density (i.e. commonality) over subsequent fixations or produced by different saccade magnitudes, along with sequential trends involving saccade angles and durations.

Applications in cognitive neuroscience

A recent trend in neuroscience has seen researchers, in addition to looking at neural responses to artificial stimuli such as bars, sinusoids and white noise, expound the value of studying neural responses to complex natural images (Barinaga, 1998). In one reported work, Vinje and Gallant (2000) achieved the realism of natural viewing by extracting image patches centred on a simulated eye movement scanpath and projecting these patches as a movie onto a neuron's receptive field. DOVES provides a valuable source of data for such research (an example collection of fixated 1° image patches are provided in Fig. 5), affording researchers the opportunity to stimulate retinal cells with a large number (30 000) fully calibrated natural stimuli, where selection for particular low level image statistics, such as dominant orientation, contrast and entropy, is also possible due to the large number and variability of fixation locations and images. Foveation may also be applied to patches in order to achieve greater realism; blurring patches in proportion to the saccade length from the (previous) fixation at which they were selected for fixation by our human observers, using functions of resolution falloff to eliminate high spatial frequencies that were not perceivable when particular patches were selected for fixation (Geisler and Perry, 1998).



Figure 5. Example fixated image patches.

Memory-related experiments

Since a visual memory task was incorporated into the eye movement recording procedure, the potential exists to compare the properties of recalled *vs.* non-recalled patches tested at the end of each trial. Differences between these groups (described below) might occur due to duality in fixation target selection: ‘centre of gravity’ (Najemnik and Geilser, 2005) or exploratory saccades designed merely to cast gaze into a new area, along with saccades drawn by particularly striking image features. Alternative explanations include that non-recalled patches represent the lower bound of a range of fixated patch saliencies, not sufficiently well represented to withstand delay prior to the recognition memory test.

Our visual task yielded a performance rate of 68%, with correct responses formed from Hit and CR responses equally. The number of yes/no responses was also equal, indicating that no response bias existed. One-way ANOVA of one common image statistic, RMS contrast, applied to patches in each response category highlights a significant effect of group, having first corrected for spurious differences between the databases used for target and lure patches by shifting the mean RMS contrast in each of the two patch groups (target and lure) to 1, $F(3, 2723) = 5.17$, $p < 0.01$. RMS contrast values for each category had homogeneity of variance, and were transformed to normal by double square root, although the application of the statistical procedures documented in this section to untransformed data does not significantly affect results (i.e. no non-significant results became significant and *vice versa*). Using protected-*t*, multiple comparisons between Hit, Miss, and FA and CR groups indicate that RMS contrast is higher for patches that resulted in corrected responses (Hit and CR) than other response categories (see Table 1). Linear contrasts formed by combining basic response categories to form yes *vs.* no, and correct *vs.* incorrect groups indicate that RMS contrast is correlated to task outcome (RMS contrast of correct response patches was found to be higher than incorrect responses, $t(2724) = 3.91$, $p < 0.01$), and that the differences observed are not attributable to response bias (RMS contrast of patches eliciting yes responses was found to be similar to no responses, $t(2724) = 0.1$, $p = 0.87$). Due to correction for differences between target and lure databases, differences in

Table 1.
Multiple comparisons across basic response categories for memory test

Comparison		<i>t</i>	<i>p</i>
Hit	Miss	2.74	<0.01**
Hit	FA	2.50	0.01**
Hit	CR	0.09	0.93
Miss	FA	-2.41	0.81
Miss	CR	-2.66	<0.01**
FA	CR	-2.41	0.02*

*Sig. at 5%, ** sig. at 1%.

RMS contrast for patches in these groups is shown to be nil (target compared to lure, $t(2724) = 0.00$, $p = 1.00$). ANOVA and subsequent multiple comparisons indicate that observers frequently fixate image regions that possess low contrast, and are unlikely to be recognised on redisplay, in addition to higher contrast regions that are likely to be recognised. Furthermore, multiple comparisons indicate that lure patches of higher contrast are more successfully rejected.

Additional analyses indicate that correct response patches are, on average, higher in contrast than general statistics of the image database they were sampled from. Recorded scan-paths were shuffled onto alternative images within the same database (thereby simulating the spatio-temporal characteristics of human eye movements, but set upon random images), revealing that only Hit rather than Miss patches possessed reliably higher than expected RMS contrast (for Hit $t(28730) = 5.18$, $p < 0.01$; for Miss, $t(28229) = 0.55$, $p = 0.58$). Tiling each of the images in the target and lure databases (165 non-overlapping tiles of 64×64 pixels each per image), calculating RMS contrast for each tile, performing the same transform as applied to the basic response categories (double square root), averaging, and then comparing to the respective RMS contrast of response categories shows that only correct responses (Hit and CR) possess higher RMS contrast than the respective tiled averages (for Hit, $t(17601) = 6.82$, $p < 0.01$; for Miss, $t(17092) = 1.27$, $p = 0.20$; for FA $t(7047) = -0.31$, $p = 0.76$, for CR $t(7513) = 3.85$, $p < 0.01$). This suggests that, of the fixations recorded, those that resulted in Miss responses possessed no special contrast characteristics when compared to the image database's general statistics, but the Hit category shows elevated RMS contrast. For signal absent response categories (FA and CR), these analyses shows that observers were particularly adept at rejecting lure patches (CR) when RMS contrast was randomly serendipitously higher than average.

Additional analyses using this feature of the data could be in relation to the eye movement statistics (such as saccade length, angle and fixation duration) that produced the patches used in 'target' trials (see the Visual task section for experimental procedure). The possibility that patches resulting in Miss *vs.* Hit responses stem from centre-of-gravity/exploration *vs.* stimulus driven saccades can be tested by relating task patches to the saccade magnitude that produced them (presumably, centre-of-gravity and exploratory saccades will be reliably longer) and fixation order (i.e. if Miss patches are simply lower in saliency, and therefore quantitatively rather than qualitatively different, they might occur later in the viewing period when all other salient areas have been exhausted). Furthermore, the database also provides a source of image patches that were each selected by human observers in a similar cognitive state, for the study of visual short term memory using isolated sequential or parallel stimuli with low/stable cognitive impact (examples are provided in Fig. 5), with the potential to match for particular low-level image properties such as luminance, contrast, entropy or orientation, or to select recalled rather than non-recalled patches from within the set of fixated image regions for assured high saliency.

Summary

In this paper, a database of high-accuracy eye movements on calibrated natural images is presented. Acquisition procedures are described, and common eye movement statistics for the captured data are calculated and summarised. The database is provided for download and use in a number of vision applications: *inter alia*, the comparison of competing gaze modelling algorithms, and the development of new algorithms, both in terms of saliency estimation, and mimicking the distributions of basic eye movement statistics; harvesting fixated image patches for electrophysiology and studies of visual memory. We outline several applications of DOVES, providing summary details of analyses relating to low level image statistics at point of gaze, eye movement statistics and RMS contrast statistics in visual memory.

Acknowledgements

This research was supported by a grant from the National Science Foundation (ITR-0427372) and by ECS-0225451.

NOTES

1. In fact, either their first fixation following the central forced fixation on stimulus onset, or the last complete fixation before stimulus offset.

REFERENCES

- Applied Science Laboratories (1998). *Eye Tracking System Instruction Manual* (Version 1.2). Applied Science Laboratories, Bedford, MA.
- Barinaga, M. (1998). Neuroscience: researchers go natural in visual studies, *Science* **282**, 614–616.
- Barlow, H. (1961). Possible principles underlying the transformation of sensory messages, in: *Sensory Communication*, Rosenblith, W. (Ed.), pp. 217–234. MIT Press, Cambridge, MA.
- Brainard, D. H. (1997). The psychophysics toolbox, *Spatial Vision* **10**, 433–436.
- Einhäuser, W. and König, P. (2003). Does luminance-contrast contribute to a saliency map for overt visual attention?, *Europ. J. Neurosci.* **17**, 1089–1097.
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells, *J. Opt. Soc. Amer. A* **4**, 2379–2394.
- Findlay, J. M. and Gilchrist, I. D. (2003). *Active Vision: The Psychology of Looking and Seeing*. Oxford University Press, Oxford, UK.
- Geisler, W. S. and Perry, J. (1998). A real-time foveated multiresolution system for low bandwidth video communication, in: *SPIE Human Vision and Electronic Imaging III*, vol. 3299, pp. 294–305. San Jose, CA, USA.
- Hancock, P. J. B., Baddeley, R. J. and Smith, L. S. (1992). The principal components of natural images, *Network: Computation in Neural Systems* **3**, 61–70.
- Hayhoe, M. and Ballard, D. (2005). Eye movements in natural behavior, *Trends in Cognitive Sciences* **9**, 188–194.

- Henderson, J. M. (2003). Human gaze control during real-world scene perception, *Trends in Cognitive Neuroscience* **7**, 498–504.
- Henderson, J. M. (2005). Introduction to real-world scene perception, *Visual Cognition* **12**, 849–851.
- Henderson, J. M. (2007). Regarding scenes, *Curr. Directions in Psychol. Res.* **16**, 219–222.
- Itti, L. and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention, *Vision Research* **40**, 1489–1506.
- Kienzle, W., Wichmann, F. A., Schölkopf, B. and Franz, M. O. (2006). A nonparametric approach to bottom-up visual saliency, in: *Advances in Neural Information Processing Systems*, vol. 19, pp. 689–696. MIT Press, Cambridge, MA, USA.
- Li, Z. (2001). A saliency map in primary visual cortex, *Trends in Cognitive Science* **6**, 9–16.
- Li, Z. and Snowden, R. (2006). A theory of a saliency map in primary visual cortex (V1) tested by psychophysics of colour-orientation interference in texture segmentation, *Visual Cognition* **14**, 911–933.
- Najemnik, J. and Geisler, W. S. (2005). Optimal eye movements strategies in visual search, *Nature* **343**, 387–391.
- Párraga, C. A., Troscianko, T. and Tolhurst, D. J. (2000). The human visual system is optimised for processing the spatial information in natural visual images, *Current Biology* **10**, 35–28.
- Parkhurst, D. J. and Niebur, E. (2003). Scene content selected by active vision, *Spatial Vision* **16**, 125–154.
- Parkhurst, D. J. and Niebur, E. (2004). Texture contrast attracts overt visual attention in natural scenes, *Europ. J. Neurosci.* **19**, 783–789.
- Parkhurst, D., Law, K. and Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention, *Vision Research* **42**, 107–123.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies, *Spatial Vision* **10**, 437–442.
- Privitera, C. and Stark, L. (2000). Algorithms for defining visual regions-of-interest: comparison with eye fixations, *IEEE Transaction Pattern Analysis and Machine Intelligence* **22**, 970–982.
- Rajashekar, U., van der Linde, I., Bovik, A. C. and Cormack, L. K. (2007). Foveated analysis and selection of visual fixations in natural scenes, *Vision Research* **47**, 3160–3172.
- Rajashekar, U., van der Linde, I., Bovik, A. C. and Cormack, L. K. (in press). GAFFE: A gaze-attentive fixation finding engine, *IEEE Transactions in Image Processing*.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research, *Psychol. Bull.* **124**, 372–422.
- Rayner, K., Li, X., Williams, C. C., Cave, K. R. and Well, A. D. (2007). Eye movements during information processing tasks: individual differences and cultural effects, *Vision Research* **47**, 2714–2726.
- Reinagel, P. and Zador, A. M. (1999). Natural scene statistics at the centre of gaze, *Network: Computation in Neural Systems* **10**, 341–350.
- Simoncelli, E. P. and Olshausen, B. A. (2001). Natural image statistics and neural representation, *Annu. Rev. Neurosci.* **24**, 1193–1216.
- Tatler, B. W., Baddeley, R. J. and Vincent, B. T. (2006). The long and the short of it: spatial statistics at fixation vary with saccade amplitude and task, *Vision Research* **46**, 1857–1862.
- Tatler, B. W. (2007). The central fixation bias in scene viewing: selecting on optimal viewing position independently of motor biases and image features, *J. Vision* **14**, 1–17.
- Torralba, A. (2003). Modeling global scene factors in attention, *J. Opt. Soc. Amer. A (Optics, Image Science and Vision)* **20**, 1407–1418.
- Torralba, A., Oliva, A., Castelano, M. S. and Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search, *Psychol. Rev.* **113**, 766–786.
- Treue, S. (2003). Visual attention: the where, what, how and why of visual saliency, *Current Opinions in Neurobiology* **13**, 428–432.

- Underwood, G. and Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures, *Quart. J. Exper. Psychol.* **59**, 1931–1949.
- Underwood, G., Foulsham, T., van Loon, E., Humphreys, L. and Bloyce, J. (2006). Eye movements during scene investigation: a test of the saliency map hypothesis, *Europ. J. Cognitive Psychol.* **18**, 321–342.
- van Hateren, J. H. and van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex, *Proc. Biol. Sci.* **265**, 359–366.
- vanRullen, R. (2005). Visual saliency and spike timing in the ventral visual pathway, *J. Physiol. (Paris)* **97**, 365–377.
- Vinje, W. E. and Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision, *Science* **287**, 1273–1276.
- Wolfe, J. M. and Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it?, *Nature Rev. Neurosci.* **5**, 495–501.
- Yarbus, A. L. (1967). *Eye Movements and Vision*. Plenum Press, New York, USA.