Perception advance online publication

doi:10.1068/p6142

Visual memory for fixated regions of natural images dissociates attraction and recognition

Ian van der Linde
Department of Computing & Technology, Anglia Ruskin University, Bishops Hall Lane, Chelmsford CM1 1SQ, UK and Center for Perceptual Systems, Department of Psychology, University of Texas at Austin, Austin, TX 78712, USA; e-mail: i.v.d.linde@anglia.ac.uk

Umesh Rajashekar¶, Alan C Bovik, Lawrence K Cormack
Center for Perceptual Systems, Department of Psychology, University of Texas at Austin, Austin, TX 78712, USA

Received 30 July 2008, in revised form 10 May 2009; published online 24 July 2009

Abstract. Recognition memory for fixated regions from briefly viewed full-screen natural images is examined. Low-level image statistics reveal that observers fixated, on average (pooled across images and observers), image regions that possessed greater visual saliency than non-fixated regions, a finding that is robust across multiple fixation indices. Recognition-memory performance indicates that, of the fixation loci tested, observers were adept at recognising those with a particular profile of image statistics; visual saliency was found to be attenuated for unrecognised loci, despite that all regions were freely fixated. Furthermore, although elevated luminance was the local image statistic found to discriminate least between human and random image locations, it was the greatest predictor of recognition-memory performance, demonstrating a dissociation between image features that draw fixations and those that support visual memory. An analysis of corresponding eye movements indicates that image regions fixated via short-distance saccades enjoyed better recognition-memory performance, alluding to a focal rather than ambient mode of processing. Recognised image regions were more likely to have originated from areas evaluated (a posteriori) to have higher fixation density, a numerical metric of local interest. Surprisingly, memory for image regions fixated later in the viewing period exhibited no recency advantage, despite (typically) also being longer in duration, a finding for which a number of explanations are posited.

1 Introduction

Shepard (1967) presented over 600 pictures sequentially to observers, who, in a subsequent old/new recognition-memory task, were able to correctly designate over 98%. Using a similar procedure, Haber (1970) found a 90% recognition rate using over 2500 pictures, and Standing (1973) found an 83% recognition rate for 10 000 pictures. These experiments, and others like them, were taken as evidence that humans possess an enduring, high-capacity scene memory. Subsequently, high performance in long-term recognition memory for pictures was proposed to occur not because our visual memory is remarkably accurate (as was originally thought), but because an imprecise, descriptive summary (or 'gist') is quickly generated and stored in lieu of precise visual detail (Irwin and Andrews 1996; Wolfe 1998). Consequently, visual memory for scenes is limited in accuracy, and can even be embellished with contextually feasible features that were not actually presented (Brewer and Treyans 1981; Intraub and Richardson 1989; Intraub et al 1992). Indeed, recent studies of change blindness have demonstrated that, quite apart from being able to retain precise visual details from large numbers of scenes viewed over an extended period, observers are scarcely able to detect dramatic changes in a single scene shown alternately with a blank interval (Simons and Levin 1997; Simons and Rensink 2005).

¶ also Laboratory for Computational Vision, New York University, New York, NY 10003, USA.

In typical studies of visual short-term memory (VSTM), the system implicated in the storage of visual stimuli over periods of several seconds, change detection is used to measure the ability of observers to detect modifications applied to a study stimulus in a subsequently displayed test. Stimuli often comprise isolated object or pattern arrays (Phillips 1974; Irwin 1991; Luck and Vogel 1997; Alvarez and Cavanagh 2004), rather than complete scenes; a recurrent finding is that observers are able to retain information pertaining to only 3 – 4 items (Irwin 1992; Cowan 2001), although, more recently, this has been shown to be contingent on informational load, with varying capacity estimates for different stimulus complexities (Alvarez and Cavanagh 2004). Interestingly, the commonly cited capacity estimate of 3 – 4 items has been shown to hold even when the stimuli to be remembered comprise different numbers of independent features, indicating that distinct feature dimensions bind to make objects, the proposed units of VSTM (Luck and Vogel 1997). Conversely, when intra-categorical stimuli are used (such as ovals of different aspect ratios), capacity falls to around one item, indicating that multi-item visual memory is reliant on our ability to uniquely categorise objects, implying copious support from longer-term semantic memory systems (Olsson and Poom 2005).

Relating visual short-term memory to viewing behaviour, it has been proposed that objects bind to locations in our visual environment (Kahneman and Treisman 1984), a notion referred to as object file theory (OFT). The fusion of visual and spatial object attributes is aptly demonstrated in experiments finding superior recognition performance where both appearance and the absolute or relative spatial positions of stimuli are preserved from study to test (Jiang et al 2000; Olson and Marshuetz 2005; Hollingworth 2006, 2007). An important extension to OFT incorporating eye movements is trans-saccadic object file theory (Irwin 1992; Irwin and Andrews 1996). Although the integration of visual information across saccades is severely limited (Irwin 1991; Henderson and Hollingworth 1999), recent studies have used eye tracking to measure observers' visual memory for objects appearing in natural scenes relative to the eye movements executed. In a change-detection framework, it has been demonstrated that object manipulations (deletions/substitutions) at previously fixated screen regions are more likely to be detected than those at non-fixated regions (Henderson et al 2003). In one novel experiment, Irwin and Zelinsky (2002) permitted a finite number of fixations to be made on a 7-object study scene comprising 7 (fixed) locations. A spatial probe was then displayed, followed by a 7-object forced-choice task, in which observers were instructed to select the particular object that had occupied the location highlighted by the preceding probe. Memory performance was found to be higher for foveated than for non-foveated objects, especially where fewer fixations were made. Accuracy was analysed in relation to how recently target objects had been foveated, yielding a pronounced recency effect: 90% accuracy was measured for the last fixated object, falling to 80% for the second and third from last and reaching a plateau at 65% for other fixations (referred to as the pre-recency level), mirroring simpler VSTM experiments in which eye gaze was not measured (above), leading researchers to propose that trans-saccadic memory and VSTM are subserved by common mechanisms. In a subsequent study Zelinsky and Loschky (2005) used a procedure in which, after observers had fixated a pre-determined target from nine arranged in a simple scene, a pre-determined number of subsequent fixations on other objects were permitted before test. At test, a spatial probe was displayed at the location previously occupied by the target, followed by a choice of four possible objects, such that one was always the target and the remaining three were other objects that had also featured in the study scene. Performance was evaluated relative to the number of intervening objects fixated following the target, prior to test. Pre-recency levels suggested relatively good scene memory irrespective of the number of intervening objects fixated (asymptote at 65%, despite a 25% chance level), a finding compatible with that of Irwin and Zelinsky (2002) and comparable studies by other researchers (Henderson and Hollingworth 2002), implying longer-term memory involvement. Furthermore, the authors proposed that eye movements effectively serialise scene-based objects, accounting for capacity and recency effects that correspond closely to sequential display conditions.

In addition to studying the formation of visual memories during scene viewing, an understanding of how observers select image regions for foveal scrutiny is necessary to gain a more complete understanding of active vision (Findlay and Gilchrist 2003). Recent studies have greatly improved our understanding of eye movements executed during the exploration of real-world scenes (Henderson 2003; Torralba et al 2006). One line of investigation has seen researchers compute natural-scene statistics directly at observers' point of gaze, and thereafter establish the degree to which the statistical properties of these image regions differ from those of regions selected randomly. For example, Reinagel and Zador (1999) found that image regions centred at human fixations have higher average spatial contrast and spatial entropy than random regions, indicating that human eye movements target image loci that maximise the information transmitted to the visual cortex. Similar findings for other image statistics (and weighted combinations thereof) were subsequently reported (Itti and Koch 2000; Privitera and Stark 2000; Parkhurst et al 2002; Einhäuser and König 2003; Parkhurst and Niebur 2003, 2004; Tatler et al 2006; Rajashekar et al 2007), complemented by work that has focused on top – down/contextual fixation guidance mechanisms (Torralba 2003; Torralba et al 2006). Top – down and bottom – up fixation guidance models are not necessarily in contention, with several researchers proposing that each may have precedence at different times, eg bottom – up effects may be of greater importance on stimulus onset where top – down knowledge is unavailable (Li and Snowden 2006), and during particular tasks, eg more so in recognition memory rather than search (Underwood and Foulsham 2006; Underwood et al 2006). Further, it has been proposed that information derived from both bottom – up and top – down mechanisms may unite in the visual stream (Treue 2003; vanRullen 2005), possibly using a saliency map that weights attraction at each location in the visual field relative to its neighbours (Koch and Ullman 1985; Li 2001). Alternative interpretations assert that bottom – up saliency is an artifact of the tendency of observers to fixate semantically important scene regions which simply co-occur with a statistical profile that suggests elevated saliency (Tatler et al 2005), finding a significant degree of correlation between the locations of human fixations on images and independent observer's ratings of local semantic interest (Mackworth and Morandi 1967; Henderson et al 2007).

Though measuring the fundamental properties of VSTM with synthetic non-categorical stimuli (such as dot-patterns or ovals) or sterile displays of organised objects are perfectly valid experimental paradigms, natural vision operates in a richly categorical environment in which overt attention is shifted through frequent eye movements, where objects vary in scale and viewpoint, incur superimposition, and in which verbal support operates unguarded. Experimental procedures deliberately suppressing verbal support, such as those used in Luck and Vogel (1997), lack full correspondence to real-world viewing conditions, as do experiments with synthetic stimuli (Phillips 1974; Luck and Vogel 1997; Alvarez and Cavanagh 2004), or neatly arranged arrays of natural or synthetic objects (Irwin and Zelinsky 2002; vanRullen and Koch 2003; Liu and Jiang 2005; Unema et al 2005; Zelinsky and Loschky 2005). Furthermore, it is likely that, in practice, visual memory in natural-scene viewing relies upon a combination of VSTM, abstract gist, and support from longer-term memory systems (Melcher 2006).

In this study, VSTM for natural scenes and gaze modeling are combined in a single experiment, with a procedure requiring observers to designate regions of viewed images as old (belonging to the image just viewed), or new. Unbeknown to observers, image regions presented at test, which were similar to the 'cut-outs' used by Velichkovsky et al (2005), were, in target trials, selected from their own fixation coordinates, enabling the effects of visual saliency and subsequent recognition to be studied in tandem. Though, in principle, similar to earlier studies measuring recognition memory for fixated objects (Henderson and Hollingworth 2002; Irwin and Zelinsky 2002; Zelinsky and Loschky 2005), in this experiment, full-screen natural images serve as stimuli, enabling observers to saccade to any image regions that drew their attention rather than one of a finite number of orderly and neatly demarcated objects. Image statistics of correctly and incorrectly identified image regions were compared with each other, but also with those of all fixated regions, and a set of pseudo-random regions created by shuffling human fixation coordinates onto alternate images.

Anatomical and behavioural evidence supports the notion that the human visual system possesses two broadly distinct cortical pathways that process the what and where/how aspects of visual stimuli, emanating from the ventral and dorsal streams of V1, respectively (Underleider and Mishkin 1982; Jeannerod and Rossetti 1993; Milner and Goodale 1995). In recent studies it has been assumed that eye-movement parameters, particularly saccade length and fixation duration, may provide a means to ascribe moment-by-moment visual behaviour to dorsal or ventral mechanisms (Velichkovsky 2002). Since the stimuli used in this study did not feature objects presented in an organised array (ie were not spatially pre-biased), the impact of eye-movement properties on recognition memory may be evaluated; these comprised fixation density, fixation duration, fixation distance from the screen centre, and preceding saccade length. In ambient mode, attributed to the dorsal pathway (with principally magnocellular input), all areas of the visual field are assumed to be represented, and thus eye movements associated with this mode may incorporate large saccades (Post et al 2003). In contrast, focal mode, attributed to the ventral pathway (with principally parvocellular input), is purported to process only a small area about the fovea, but to enjoy access to visual and semantic memory (Creem and Proffitt 1999); should this theory be correct, saccade length and recognition memory should be demonstrably correlated, even after correcting for the possibility that multiple small saccades may generate re-fixations of the same object/region.

2 Method

2.1 *Apparatus*

Observers' eye movements were recorded with an SRI/Fourward Generation V Dual Purkinje eye tracker (Crane and Steele 1985) with a spatial accuracy of < 10 min of arc, and a response time of ~ 1 ms. A bite-bar and forehead-rest were used to minimise head movements. Monocular tracking was used to reduce calibration time. The output of the eye tracker was sampled at 200 Hz and stored for off-line analysis. Images were viewed on a 21 inch grey-scale monitor (Image Systems Corp., Minnetonka, MN) placed at a distance of 134 cm, running at refresh rate 60 Hz with screen resolution of 1024×768 pixels. The monitor was calibrated, in that the gamma function was linearised prior to use. This ensured veridical display of the calibrated images used (see section 2.3), in that the luminance relationships between pixels established during acquisition (with a calibrated camera) were accurately reproduced at display.

2.2 *Participants*

Twenty-nine unpaid adult human volunteers (nineteen male, ten female, mean age 27 years, SD 5.55 years) served as observers, each with normal or corrected-to-normal vision.

2.3 *Stimuli*

A set of 101 calibrated grey-scale natural images (van Hateren and van der Schaaf 1998) served as stimuli. Images were composed of a few natural features (trees, grasses, bushes, soil, sky, and water) which, in metrics like valence and arousal (Maljkovic and Martini 2005), were considered to be roughly equal. We omitted images containing both man-made structures and features such as humans, animals, and other content of particularly obvious high-level cognitive interest that would have instinctively attracted attention (Buswell 1935; Yarbus 1967). Example images, with overlaid scan path and fixations for a single observer, are shown in figure 1. The images from the database have a native resolution of 1536 × 1024 pixels. A central region of 1024 × 768 pixels was cropped from within each image to fit the desired screen configuration. Images were luminance-scaled such that the brightest pixel in the image corresponded to the brightest output level of the monitor. In addition to full-screen study images, 1 deg × 1 deg (64 × 64 pixel) test patches were displayed (see section 2.4). Stimulus presentation was controlled in Matlab with the Psychophysics Toolbox (Brainard 1997; Pelli 1997).

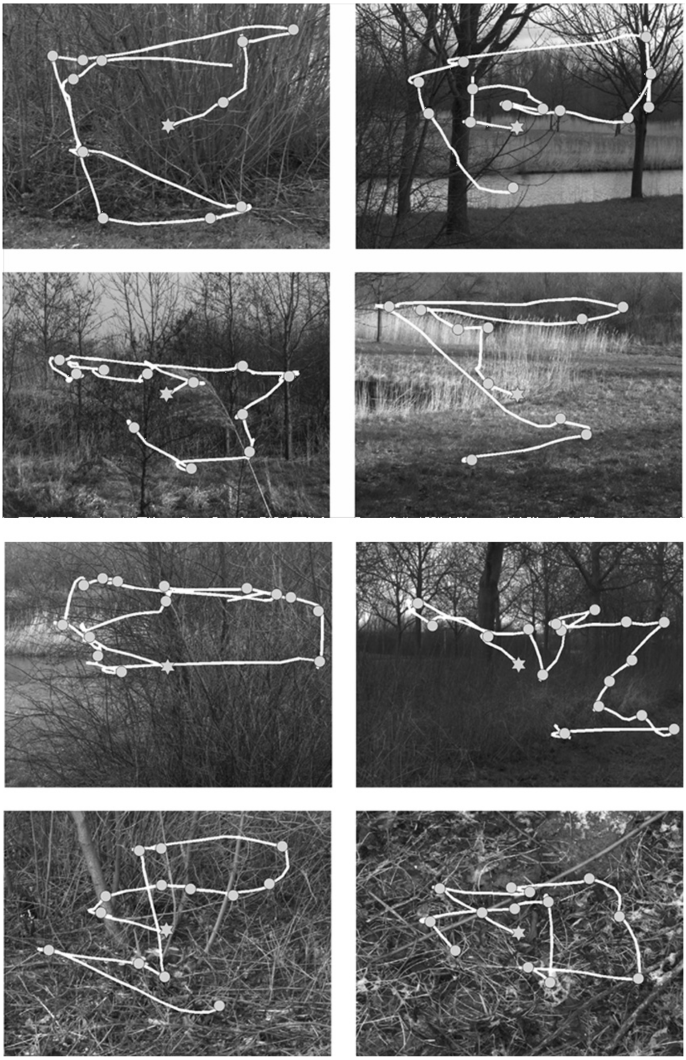

Figure 1. Example images with overlaid eye movements (serpentine white lines), and fixations (grey circles) for one observer. The start position (at the screen centre) is denoted by a star symbol.

2.4 *Procedure*

101 trials, in which observers viewed the 101 stimuli (above), were completed by all twenty-nine observers, yielding a total of 2929 trials. For each observer, a 3×3 calibration grid was shown prior to a block of (maximum length) 10 trials. A calibration test was performed before each trial, requiring observers to fixate a 0.3 deg central marker within 5 s. In the event of failure of the calibration test, the full calibration routine was repeated. A central fixation test before each trial ensured that all observers commenced viewing the stimuli from the same location, standardising the viewing experience across trials and observers. On passing the calibration test, a full-screen natural image was displayed for 5 s. Idiosyncratic differences in viewing behaviour will have led observers to execute different numbers of fixations within this period, including the possibility of neglecting to visit one or more image features/regions; fixations of different durations and saccades of different lengths will also have occurred.

At the end of the 5 s viewing period, a recognition memory test was administered: a 1 deg × 1 deg square (64×64 pixel) test patch was displayed at the screen centre immediately after the full-screen natural image was offset. Observers responded to indicate if they believed the test patch had originated from the image just viewed ('old'), or a different image ('new'), using a handheld keypad. During target trials, the test patch contained a region of the study image selected from one of the observers' fixation coordinates (the first free fixation, following the central calibration test fixation, or the last complete fixation). Fixation coordinates were established with a variant of an Applied Science Laboratories (1998) algorithm that dealt with drifts, blinks, and microsaccades. Displaying the test patch at the screen centre ensured that, even though no ISI was used, an eye movement was required, thereby disabling iconic memory (Sperling 1960). During lure trials, the test patch was simply selected from one of a set of 40 different images selected from the same database, satisfying the same selection criteria; as with target trials, the test patch was sampled at the observer's first free or last complete fixation coordinate (simply taken from a different image). The probability of a target or lure trial was 50%, randomised on a trial-by-trial basis. The luminance of test patches was shifted by a random value (respecting interpixel linearity, ie concatenation of distinct luminance bins was not done) to disenable the use of first-order statistical cues only (ie global brightness matching). Example test patches are shown in figure 2, which typically contained textural exemplars of flora rather than complete objects (ie the visual task typically required a relatively difficult within-category distinction between plant matter types/textures). The size and central relocation of test patches eliminated both context and positional information. Although it may have been that, on some occasions, lure patches represented objects of a type not present in the full-screen study image, it is important to note that target trial patches are the principal source of analysis (ie those in which the test patch always originated from the observers' own fixation coordinates). In target trials, the outcome of the recognition task is therefore attributable to the image properties at fixation loci visited, and eye movements executed, during free exploration of the scene.

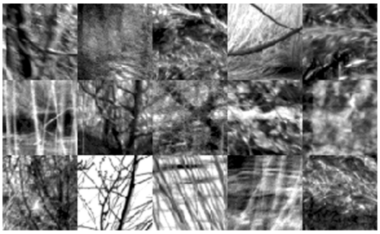

Figure 2. Example test patches used in recognition memory task.

Furthermore, considerable numbers of both miss and false alarm (FA) responses were forthcoming (see section 3), highlighting the relative difficulty of the task, notwithstanding any intermittent verbal object-identification/elimination strategy.

Participants were informed of the experimental procedure prior to their session, including the prior probability of a random response being correct; however, observers were not informed that test patches were to be sampled at their fixation coordinates. A post-experiment debriefing session revealed that none of the naive participants had become aware of this, eliminating the possibility of 'cheating' by consciously electing to keep the eyes still (see section 3). This was verified by plotting and reviewing observer's scan-paths at the end of each session.

2.5 *Analysis method*

A group of 1 deg × 1 deg image patches centred at all human fixation coordinates was collected. A group of image-shuffled fixation patches was also collected (human fixations placed upon images that were shuffled in order). The five image statistics defined in table 1 [adapted from Gonzales et al (2004)] were calculated for each group; specifically, these are modifications of the statxture function from the DIPUM library, a commonly used battery of image texture matrics.[1] Patches used in the visual-memory task were grouped according to trial type (target/lure), patch origin (first free fixation or last complete fixation), observer response (old/new), and thereafter subject to the same image statistics. Owing to the typically non-normal distributions of image statistics (Baddeley 1996), the non-parametric Wilcoxon rank sum test (Wilcoxon 1945) was used to compare between groups.

Table 1. Image statistics.

$m = \sum_{i=0}^{L-1} z_i p(z_i)$	Average luminance: higher value indicates greater patch luminance (brightness).	(1)
$\sigma = \sqrt{\mu_2(z)}$	Average contrast (standard deviation): higher value indicates greater patch contrast.	(2)
$\lvert\mu_3\rvert = \sqrt{\sum_{i=0}^{L-1} (z_i - m)^3 p(z_i)}$	Root absolute third moment: skewness of patches' intensity histogram, 0 indicating symmetry, and a positive value indicating skewness (in either direction) ie uneven utilisation of the available intensity range.	(3)
$U = \sum_{i=0}^{L-1} p^2(z_i)$	Uniformity: 1 where pixel intensities in patch are identical, falling to 0 where pixel intensities are maximally separated.	(4)
$e = \sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i)$	Spatial entropy: higher where pixel intensities in patch are randomly organised, lower when image intensities are more orderly.	(5)

Four eye-movement properties were calculated for all human fixations; the same eye-movement properties were calculated for fixations used to generate target trial patches (ie those that became hit and miss groups). These were: (1) fixation density (number of other fixations on the same image, from all observers, within a square 1.5 deg × 1.5 deg area surrounding the fixation under consideration; for specific analyses, re-fixations from

[1] Modifications from Gonzales et al (2004) included the following: smoothness was not used, since it was found to be collinear with average contrast; third moment was made absolute in order that it provided a measure of luminance skew independently of skew direction and square rooted to promote normality, thereby increasing discrimination at the distribution's centre.

the same observer within this area were excluded), (2) fixation duration (s), (3) fixation distance from screen centre (deg), (4) length of preceding saccade (deg). The Wilcoxon rank sum test was again used to compare between groups, owing to non-normal distributions of the eye movement statistics used. In analyses of both image statistics and eye movements, Bonferroni correction was used to counteract type I error rate inflation due to multiple comparisons (ie H_1 was declared at α/N, where N is the number of tests to be completed).

3 Results

3.1 *Behavioural*

Across observers, a mean correct response rate of 0.68 ($\mathrm{SD} = 0.07$) was seen, with $d' \simeq 1$ (0.92) and $\beta = 0.97$; patches were discriminable (when calculating SD across observer's individual performance rates, 68% performance is > 2.58 SDs above the 50% chance response rate, ie has a $< 1\%$ probability of having occurred by chance), with no significant response bias. Correct responses comprised nearly equal numbers of hit and correct rejection (CR) [hit-rate mean $= 0.68$, $\mathrm{SD} = 0.10$, CR-rate mean $= 0.66$, $\mathrm{SD} = 0.14$], and thus incorrect responses comprised nearly equal numbers of miss and FA. In target trials, mean hit rates for test patches derived from the first free fixation and last fixation were 0.70 ($\mathrm{SD} = 0.13$) and 0.67 ($\mathrm{SD} = 0.13$), respectively. However, a paired t-test indicates that the hit rates derived from first free versus last fixation groups were not significantly different ($t_{28} = 1.21$, $p = 0.24$). Explanations for this counterintuitive result (ie the absence of a significant recency advantage) are proposed later. An average of 12.12 fixations per observer/image were seen ($\mathrm{SD} = 3.03$), ie just under 3 fixations s^{-1}, a number consistent with seminal eye-tracking work with natural images (Buswell 1935; Yarbus 1967), indicating that images were indeed freely viewed. A mean of 12.07 fixations per image/observer in hit trials was seen ($\mathrm{SD} = 2.98$), with a mean of 12.39 ($\mathrm{SD} = 2.77$) in miss trials. Corresponding distributions are shown in figure 3a. This suggests that the number of fixations influenced the task outcome little; however, this is likely to have been because the average number of fixations executed during the 5 s viewing period enabled the images to be scanned fairly comprehensively; cf Loftus (1972), in which it was noted that memory performance was impaired where fewer fixations were permitted during a recognition test for full-screen natural scenes.

3.2 *Image statistics*

All five image statistics applied were found to be statistically significant in distinguishing human from image-shuffled fixations (Wilcoxon rank sum test results provided in table 2). Distributions of these statistics are provided in figure 4 (human: grey bars, image-shuffled: grey line). Furthermore, it was found that human and image-shuffled fixations were distinguishable in four of the five image statistics used over multiple fixation indices, excluding the first fixation which was compulsorily at the image centre for both human and image-shuffled groups (figure 5). Luminance yielded both the weakest overall discrimination (the Z value from the Wilcoxon rank sum test was $3-4$ times lower than the other image statistics), and was the only one of the five image statistics to be insufficiently divergent between human and image-shuffled human fixation groups to provide a consistent, significant difference over multiple fixation indices (figure 5a).

In the recognition-memory test administered after each trial, as above, all five image statistics were found to be statistically significant in distinguishing hit and miss response groups (table 3). Luminance was the strongest discriminator between hit and miss response groups (ie highest Z value), despite being the weakest discriminator between human and image-shuffled fixations (lowest z value, above); ie, simply stated,

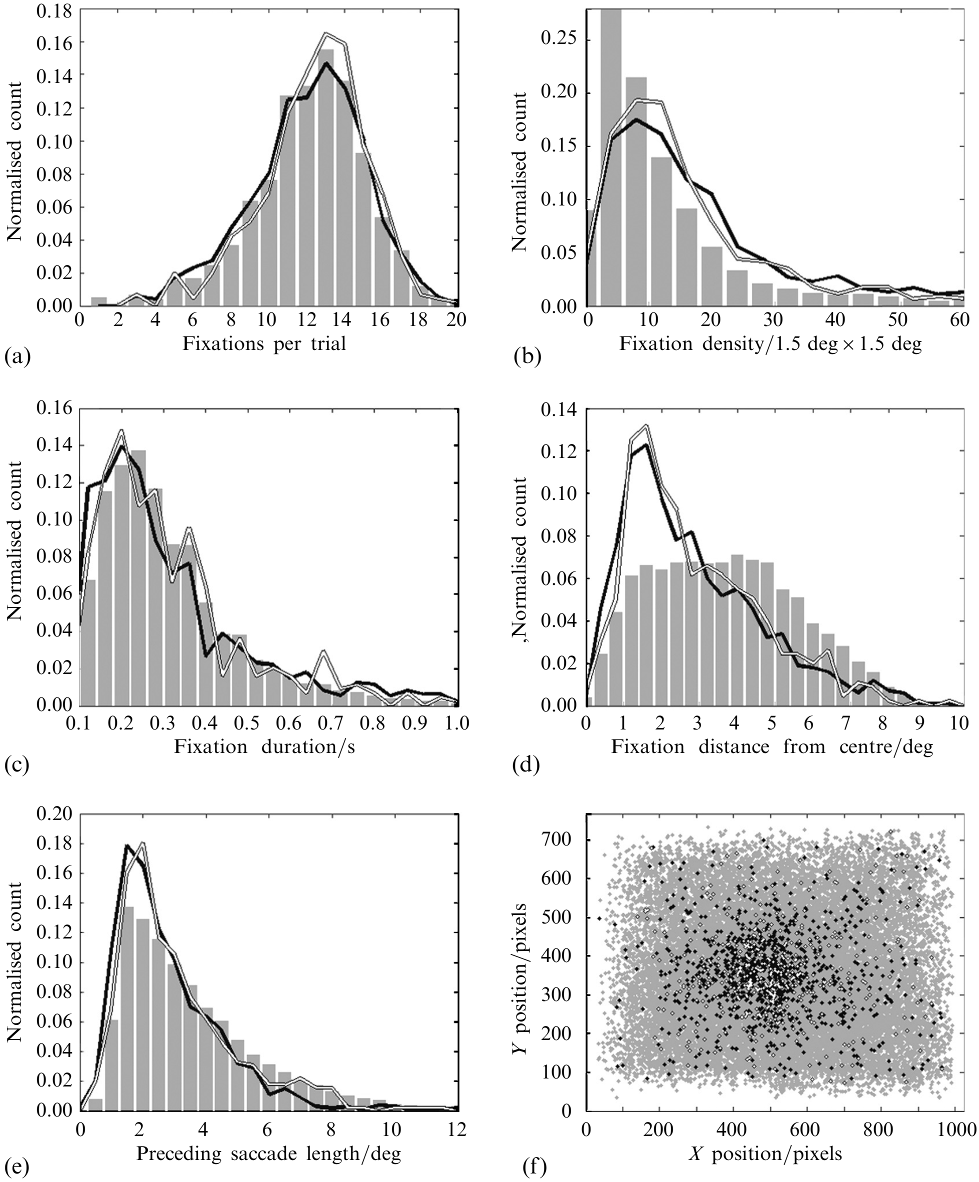

Figure 3. (a) Histogram of total fixation per trial. (b) – (e) Eye movement histograms. Grey bars: all human fixations; black line: hit group; white line: miss group. (f) Fixation coordinates (grey dots: all human fixations; black dots: hit group; white dots: miss group).

Table 2. Rank sum test comparing image statistics for human and image-shuffled human fixations. Note: the asterisk indicates significance at a probability of 95% or better.

Statistic	p	Z
m	0.00	2.96*
σ	0.00	12.02*
$\|\mu_3\|$	0.00	12.50*
U	0.00	−9.20*
e	0.00	9.90*

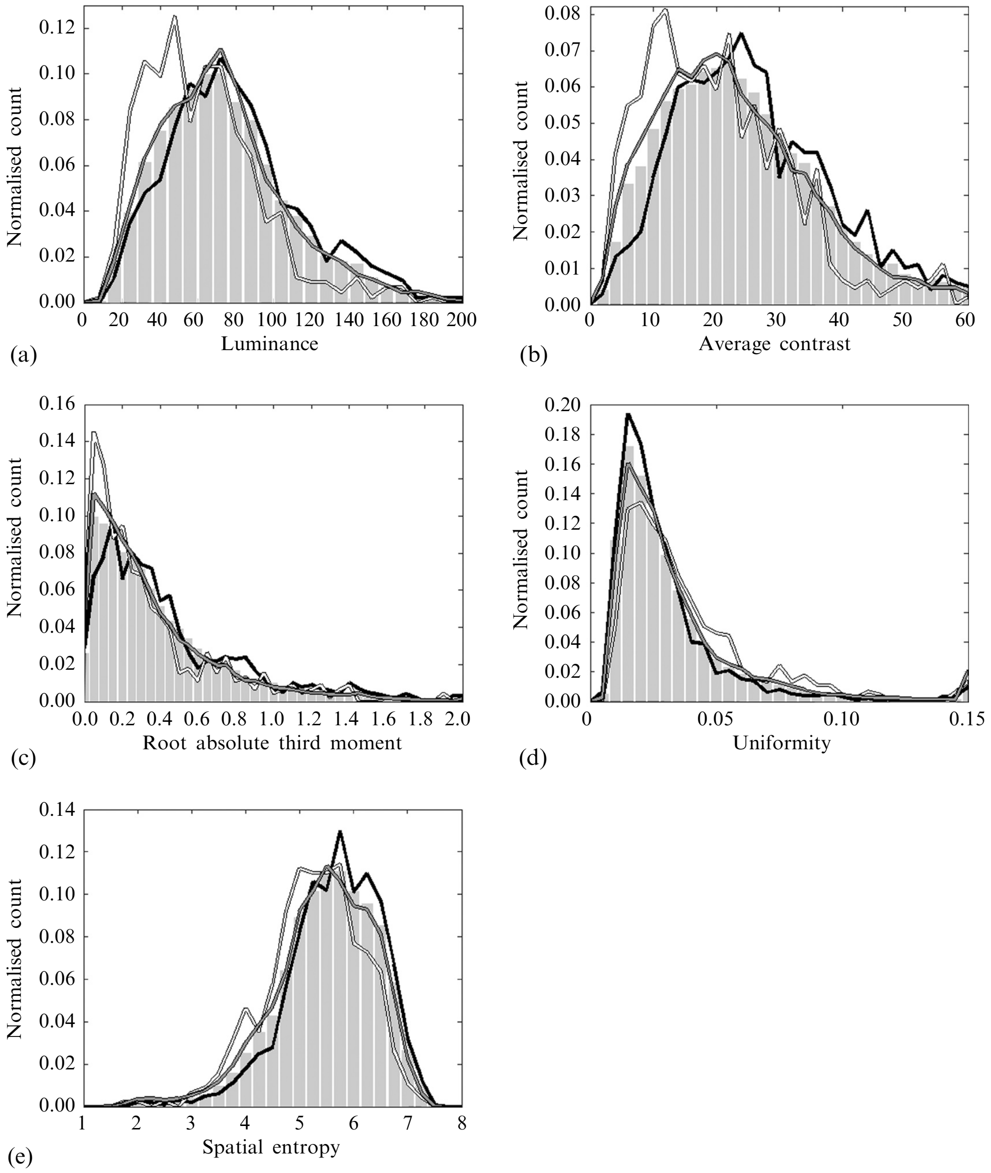

Figure 4. Histograms of local image statistics. Grey bars: all human fixations; grey line: shuffled human fixations; black line: hit group; white line: miss group.

Table 3. Rank sum test comparing image statistics for hit and miss task response groups. Note: the asterisk indicates significance at a probability of 95% or better.

Statistic	p	Z
m	0.00	7.92*
σ	0.00	7.26*
$\lvert\mu_3\rvert$	0.00	5.23*
U	0.00	−7.77*
e	0.00	7.78*

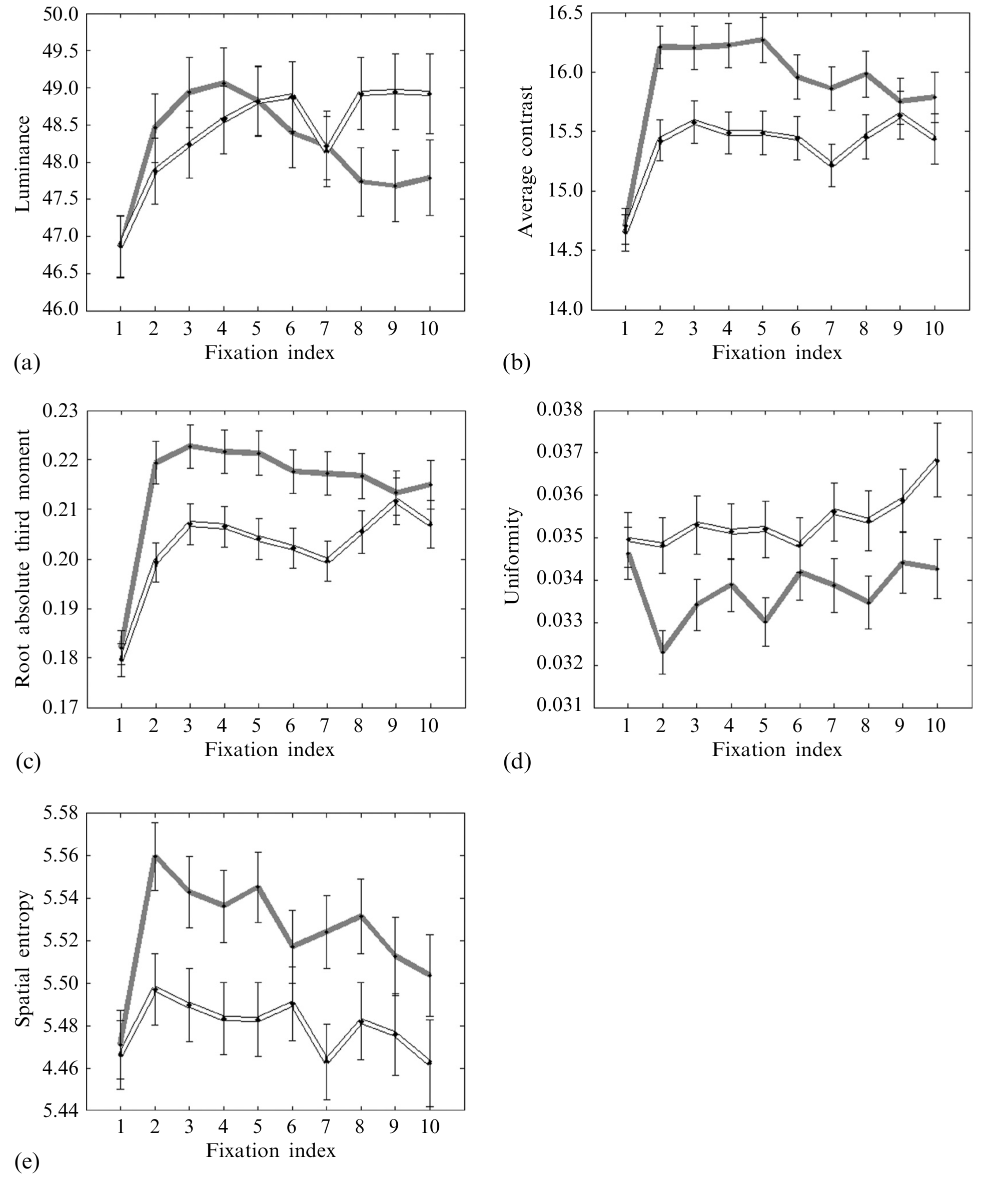

Figure 5. Mean image statistic value over sequential fixation indices. Grey line: human fixations; white line: shuffled human fixations. Error bars show ± 1 SE of the mean.

although elevated luminance was found to contribute least to drawing fixations, it was found to have the greatest influence in determining if fixated regions were subsequently recognised. The distributions of the image statistics for the hit and miss response patch groups are provided in the context of the human and image-shuffled human fixations (figure 4): in most cases, the miss distribution contains a greater number of occurrences of lower values (indicative of attenuated visual saliency) compared with the corresponding hit distribution (in uniformity this relationship is inverted, since higher value indicates greater uniformity).

A correlation between the global luminance of each image and hit rate for that image was observed ($r^2 = 0.19$, $p < 0.01$), with other image statistics, except for third moment ($r^2 = 0.08$, $p < 0.01$), a luminance-related skewness statistic, showing no significant correlation. Although this indicates that specific global image statistics affected recognition performance, possibly because of slower information acquisition rate for darker images (Loftus 1985), the majority of the variance (81%) was not attributable to the global luminance of individual full-screen stimuli, but (presumably) due to local image features at the point of gaze and associated eye-movement properties.

Given that data originating from lure trials cannot be compared to the set of all human fixations or image-shuffled human fixations, or be analysed in terms of eye movements, subsequent analyses focus on target trials: it is sufficient to observe that on a significant number of occasions lure-trial test patches enticed observers into making an incorrect response (FA), and that, during target trials, not all fixated image regions were correctly recognised (ie resulted in a miss).

3.3 *Eye movements*

Two of four eye-movement metrics applied (fixation duration, and fixation distance from centre) did not differ significantly for the patches in the hit and miss groups (table 4), helping to eliminate the possibility that miss responses were systematically shorter in duration (eg truncated at the end of the viewing period), or were more likely to have originated from particular screen locations relative to the central start position. The observation that the distance from the screen centre did not influence task performance corroborates the behavioural data indicating that the temporal source of the task patch (first or last fixation) did not significantly affect recognition rate, since the first free fixation will have tended to be closer to the screen centre owing to the compulsory central calibration test preceding each trial, and the intrinsically biased probability distribution of saccade magnitudes favouring shorter saccades (van der Linde et al 2009), among other issues (Tatler 2007). An overall central fixation bias in this study is demonstrated by a large significant inverse correlation between fixation density and fixation distance from centre ($r^2 = 0.35$, $p < 0.01$).

Table 4. Rank sum test comparing eye-movement statistics for hit and miss task response groups. (1) fixation density, (2) fixation duration, (3) distance from centre, (4) preceding saccade length. Note: the asterisk indicates significance at a probability of 95% or better; ~ indicates marginal significance.

Statistic	p	Z
1	0.03	−2.13 ~
2	0.42	0.80
3	0.45	0.75
4	0.01	2.68 *

Conversely, both preceding saccade length and fixation density were found to differ between hit and miss groups: shorter average preceding saccade length (hit mean 2.79 deg, miss mean 3.09 deg) coupled with higher fixation density for the hit group (hit mean 17.49, miss mean 15.91) were noted, suggesting that local scrutiny of a more frequently visited region was more likely to yield a hit response. Note: despite that short saccades are likely to have led to higher fixation density, confounding these variables to some degree, the correlation between them is only $r^2 = 0.04$, $p < 0.01$, and $r^2 = 0.07$, $p < 0.01$ for hits and misses respectively; and $r^2 = 0.03$, $p < 0.01$ for all fixations. Furthermore, the finding that shorter preceding saccades occur more frequently in hit rather than miss responses survives elimination of trials in which observers re-fixated within a 1.5 deg radius of the centre of a test patch (Wilcoxon rank sum test: $Z = 2.30$, $p = 0.02$), which discards 20.28% of trials. However, fixation density between hit and miss groups did not retain a significant difference after elimination of re-fixations ($Z = -1.43$, $p = 0.12$), perhaps suggesting that loci of interest were often individual to observers. Distributions of eye-movement statistics for hit, miss, and all human fixations are provided in figures 3b–3e, with a 2-D visualisation of fixation locations in figure 3f.

3.4 *Comparison of human and image-shuffled human fixations*

All five image statistics applied were found to be statistically significant in distinguishing the hit response group from all human fixations (table 5a) and image-shuffled human fixations (table 5b). Not surprisingly (given that hit patches originate from human fixations), the hit response group was found to be more dissimilar from the image-shuffled human fixation group than from the human fixation group. Luminance was found to discriminate hits from image-shuffled human fixations the least, and to be among the two least discriminatory image statistics when compared with the human fixation group (along with root absolute third moment, a luminance-related skewness statistic). Likewise, all five image statistics applied were found to be statistically significant in distinguishing the miss response group from all human fixations (table 5c) and image-shuffled human fixations (table 5d). In contrast to the hit group, the miss group's image statistics suggested a paucity of visual saliency relative to both human and image-shuffled human fixations. Interestingly, the miss group was found to be more similar to the image-shuffled human fixation group than the human fixation group of which it is a subset (exemplified by lower-magnitude Z values in the former comparison in all five image statistics), even though a significant difference overall between human and shuffled-human fixations exists (table 2). The greatest discriminator between the miss response group and both the human and image-shuffled human fixation groups was luminance, which was found to be particularly attenuated. That the miss group, generated by using fixations executed during free visual scanning, possessed lower visual salience even than the image-shuffled human fixation group may suggest self-selection of image regions that were particularly dark and ostensibly featureless (see later).

Table 5. Rank sum test comparing hit and miss groups to human and image-shuffled human fixation groups. (a) Hit versus human, (b) hit versus image-shuffled human, (c) miss versus human, (d) miss versus image-shuffled human. Note: the asterisk indicates significance at a probability of 95% or better; ~ indicates marginal significance.

Statistic	p	Z	Statistic	p	Z
(a) *Hit versus human*			(b) *Hit versus image-shuffled human*		
m	0.00	3.91*	m	0.00	4.78*
σ	0.00	4.19*	σ	0.00	7.26*
$\lvert\mu_3\rvert$	0.00	3.37*	$\lvert\mu_3\rvert$	0.00	6.50*
U	0.00	−4.05*	U	0.00	−6.41*
e	0.00	4.23*	e	0.00	6.76*
(c) *Miss versus human*			(d) *Miss versus image-shuffled human*		
m	0.00	−6.98*	m	0.00	−6.67*
σ	0.00	−5.80*	σ	0.00	−3.63*
$\lvert\mu_3\rvert$	0.00	−4.05*	$\lvert\mu_3\rvert$	0.07	4.00~
U	0.00	6.43*	U	0.00	4.78*
e	0.00	−6.33*	e	0.00	−4.55*

3.5 *Temporal order effects*

Ignoring the memory task, no significant differences between the image statistics of task patches originating from the first free fixation versus last complete fixation were found (table 6a), indicating that the image statistics applied did not change discernibly over sequential fixations, eg decreasing saliency over the viewing period as higher-saliency areas are gradually exhausted (Parkhurst et al 2002).

Table 6. Rank sum test comparing fixations originating from first free versus last fixation: (a) image statistics; (b) eye-movement statistics: (1) fixation density, (2) fixation duration, (3) fixation distance from centre, (4) length of preceding saccade. Note: the asterisk indicates significance at a probability of 95% or better.

Statistic	p	Z	Statistic	p	Z
(a) *Image statistics*			(b) *Eye-movement statistics*		
m	0.63	−0.48	1	0.00	−12.28*
σ	0.11	−1.60	2	0.00	3.53*
$\|\mu_3\|$	0.07	−1.79	3	0.00	17.55*
U	0.48	0.69	4	0.00	15.30*
e	0.34	−0.96			

Despite this, differences in eye-movement statistics were apparent. In a comparison of first free versus last fixation groups (table 6b), last fixations were seen to have longer average duration than the first free fixation (first median 0.26 s, last median 0.29 s), consistent with the tendency for fixation duration to increase over viewing period found in many tasks and studies, including those of Buswell (1935) and Yarbus (1967), but more recently Unema et al (2005) and Irwin and Zelinsky (2002); this is despite that, in this study, recognition performance for the last fixation was not found to be significantly higher than for the first free fixation. Furthermore, from first free to last fixation, a significant increase in distance from centre (first median 2.26 deg, last median 3.40 deg), along with a significant increase in preceding saccade length (first median 2.26 deg, last median 2.87 deg) and a decrease in fixation density (first mean 16.51, last mean 15.60) were found. However, density reduction, distance from centre increase, and preceding saccade increase are expected, given that final fixations do not immediately follow the forced central calibration test fixation (see earlier).

4 Discussion

In agreement with earlier studies, we find that image regions centred at human fixations are significantly different from random image regions (in this study, human fixations shuffled onto alternate images), and, furthermore that these differences are robust across multiple fixation indices (table 2, figure 5). The observation that, among the five image statistics used, luminance is least effective at distinguishing human from image-shuffled fixations is likely to be because absolute luminance is rather less important than contextual luminance in attracting fixations; ie visual saliency is higher for image regions that differ from their surroundings (in bright image areas, low luminance regions would have higher saliency, and vice versa), exemplified by good performance with the use of centre–surround mechanisms in fixation prediction algorithms (Itti and Koch 2001; Rajashekar et al 2007).

The image statistics of correctly/incorrectly recognised image regions (hits and misses), which are subgroups of the set of all human fixations, also differed significantly, with correctly identified regions possessing image statistics alluding to greater local visual saliency (table 3). Furthermore, the profile of image statistics that supported visual short-term memory was found to be distinct from the corresponding profile that drew fixations. Specifically, high luminance was found to be the most significant local-image feature distinguishing recognised and unrecognised test patches, despite being the least significant local-image feature distinguishing human from shuffled-image human fixations generally—one of the principal findings of this study, since it dissociates the image properties that draw fixations from those that support visual short-term memory, indicating that observers are frequently drawn to regions in natural images that are ultimately unmemorable.

Next, hit and miss groups were compared to the group of all human fixations, of which they are a subset, and the group of image-shuffled human fixations (table 5). Test patches leading to hit responses were found to occupy the high end of each visual saliency distribution, being significantly different in all image statistics applied (table 5a). Therefore, given that human and shuffled-image human fixation groups differ overall (table 2), it is unsurprising that an even greater discrimination between the hit and shuffled-image human fixation groups is seen (table 5b). However, a different pattern emerges in comparing the miss group to the set of all human and shuffled-image human fixations. Patches leading to miss responses were attenuated in all statistics compared to the set of all human fixations (table 5c), and in four of the five image statistics compared to the shuffled-image fixations, with the fifth statistic also being marginally significant (table 5d); indeed, the miss patch group was more similar to the shuffled-image human fixation group, despite being produced by human fixations. Patches leading to miss responses had attenuated values in all statistics applied, suggesting that, despite originating from image loci that humans freely fixated, they exhibited a paucity of visual saliency that stymied subsequent recognition. This suggests that a proportion of the fixations recorded were not well guided by local low-level image statistics, but perhaps served to cast gaze into a new area, ie possibly in favour of ambient rather than focal mechanisms, a hypothesis supported by the analysis of corresponding eye movements (see later).

The finding that luminance affects the accuracy of recognition memory is compatible with the paper by Loftus (1985), in which the information acquisition rate in picture viewing was proposed to be related to image luminance, demonstrated in both short-term and long-term memory tests, finding that extended viewing time can compensate for lower scene luminance, indicating tardy acquisition rather than poor discriminability of scene features per se. This study demonstrates that luminance affects the ability to recognise localised image regions centred at human fixations as well as entire images (as demonstrated by Loftus), even though the specific image patches tested were freely fixated by observers. Alternative explanations for poor recognition performance for test patches generated from lower luminance fixations include that intrinsic noise accumulates more quickly as signal properties (such as luminance) are proportionally reduced (Sperling 1986).

Ssaccade length was found to be shorter, and fixation density higher, in the hit group (table 4); after elimination of re-fixations within a 1.5 deg × 1.5 deg area, saccade length remains significantly shorter in the hit response group. The observation that fixations produced by shorter saccades are more conducive to subsequent recognition is compatible with the findings of Velichkovsky et al (2005), who proposed that shorter saccades are furnished with access to the memory systems supporting the ventral processing stream. Although it is important to note that, ultimately, all fixated regions gain access to the ventral stream and the memory support that this pathway purportedly affords, one explanation for the reduced recognition performance for fixations produced by lengthy saccades is that they may have been less likely to have been selected after semantic identification (being represented primarily by low-spatial-resolution magnocellar systems), potentially contributing to the paucity of striking local-image features at the saccade target, especially, as some researchers contend, if visual saliency and semantically interesting scene regions simply co-occur. However, an important caveat with respect to the analysis of saccade length should be noted: notwithstanding the 1.5 deg × 1.5 deg exclusion zone to remove re-fixations from our analyses, it may be that superior performance for test patches derived from shorter saccades originates from greater extrafoveal perception (ie previous fixations occurring adjacent to the test patch); the exclusion of re-fixations goes some way to mitigating this possibility, but it cannot, nevertheless, be entirely discounted. Furthermore, previous work has emphasised the importance of saccade length in the consideration of visual saliency, indicating the short distance saccades are more likely to be directed to image regions possessing greater visual saliency (Tatler et al 2006), which is consistent with an analysis of the fixation data used in this study (Rajashekar et al 2007). If shorter saccades tend to be directed to more visually salient image regions, greater memory performance for image regions generated via short saccades may actually be the result of that elevated saliency rather than that short saccades accrue support from ventral memory systems. In practice, it may be that these confounded factors can only be satisfactorily extricated by using neuroimaging to identify whether activation in the dorsal/ventral pathways is modulated as a function of saccade length and other eye-movement parameters.

Although behavioural performance data indicate that observers were able to distinguish target from lure patches at significantly greater than chance level, with hit patches typically possessing greater visual saliency originating from more-frequently visited locations (ie higher a posterior fixation density), and being produced by shorter saccades, test patches produced by first free and last complete fixations did not produce significantly different performance rates. Thus, unlike in the studies of Irwin and Zelinsky (2002) and Zelinsky and Loschky (2005), but like in that of Henderson and Hollingworth (2002), no recency effect was apparent. A number of possibilities exist why this was the case. One possibility is that a tendency for observers to select image regions with especially striking visual properties earlier in the viewing period may have existed (Parkhurst et al 2002), which consequently accrue a more enduring representation in memory; however, a comparison of first free and last fixations without taking task responses into account (table 6a) shows no significant decrease in any image statistic measured, making this explanation improbable. Other possibilities include that earlier-fixated regions were able to migrate to long-term memory (~5 s), or that semantically incongruent (and thus more memorable) scene features were fixated earlier in the viewing period (Henderson et al 1999) and are not captured by the image statistics applied. Like Henderson and Hollingworth (2002), we quantify recency using fixation indices rather than object foveations, which may dilute recency measurements (Zelinsky and Loschky 2005), but given the ~5 s intermission between observers' first and last fixation, this is unlikely to have been a sufficiently strong factor to entirely eliminate recency effects. However, it is reasonable to assume that several fixations are used to scrutinise single objects, and thus visual short-term memory object capacity may have not been saturated despite an average of 12 fixations having been made (however, in terms of recency, significant differences in image statistics survive the elimination of re-fixations, making this interpretation also rather unlikely). Further to this, because the natural stimuli used feature objects that were large in scale and repetitive in appearance, significant potential for extrafoveal support may have existed, stemming the decay of earlier fixated regions. However, it is unlikely that the absence of a verbal suppression task was the primary cause for our findings, given the striking differences between the image statistics of the hit and miss groups. Paraphrasing the argument of Irwin and Zelinsky (2002), who also use a 5 s viewing period, verbal memory supports 5–9 items and it is implicit that, had this been the principal strategy (ie the retention of object names for later identification), a rather higher performance rate would have been expected than that observed. Similar reasoning can be used to propose that gist retention was not overarching strategies (insofar that this strategy is even possible, given the relative homogeneity of the stimuli used in this study), since this would, likewise, not have produced significant differences in either eye movement or image statistics between hit and miss groups.

The fixation of image regions that are darker and less texturally varied even that the shuffled-image human fixation group suggests self-selection of relatively featureless loci. Aside from interpretations based on the use of centre-of-gravity/exploratory saccades, higher saliency for dark regions in bright images, correlation between saccade length and saliency, or alternation between dorsal/ventral mechanisms that entail dissimilar memory support, a rather more speculative interpretation could be that a proportion of human fixations are dedicated to the investigation of dim image regions of natural images that could conceal predators or other perils as part of an instinctive/evolved viewing behaviour, or, more simply, that relatively featureless image regions may actually possess high saliency (to human observers) when appearing in an image containing a great deal of visual disorder. However, unrecognised test patches were also produced by image regions with lower fixation density and shorter preceding saccade length, which cannot be satisfactorily accounted for by such mechanisms.

The proposal by some researchers that human fixations co-occur with elevated visual saliency simply because local saliency and local semantic interest are correlated (Mackworth and Morandi 1967; Henderson et al 2007) cannot explain why, in this study, a significant proportion of fixations were found to be placed at image loci that had lower saliency than even random regions that, subsequently, observers were unable to recognise. Our findings support the notion that not all fixations are driven by striking local attractors; indeed, this may account for the often relatively weak statistical separation between human and random fixations in many published studies, since, in such a scenario, a proportion of the population of all human fixations will be given over to (ostensibly) low-saliency image regions (indeed, in figure 4, showing image statistics of human and shuffled-image fixations in this study, the differences between hit and miss distributions is very much greater than the difference between human and shuffled-image human fixations overall).

5 Conclusions

Taken together, our results suggest a duality in selecting fixation loci: observers were seen to fixate image regions the statistics of which suggest high visual saliency, which were subsequently correctly recognised, in addition to image regions whose statistics suggested attenuated saliency, which observers were subsequently unable to recognise. Less-salient regions may correspond to centre-of-gravity or exploratory fixations designed to cast gaze into a new area, rather than being drawn by striking image features or focal object scrutiny. Supporting this hypothesis, we find that unrecognised regions were more likely to be produced by longer saccades, and, a posteriori, were found to have lower fixation density. Alternatively, unrecognised regions may represent fixation loci at the lower bound of the image statistics measured (being significantly beneath corresponding values from the set of all human fixations), and may either decay prior to the recognition-memory test, owing to, for example, faster accumulation of intrinsic noise (Sperling 1986), or slower initial encoding (Loftus 1985).

The finding that successfully recognised image regions are more likely to have been produced by shorter saccades was also noted by Velichkovsky et al (2005) in a study designed to measure memory for fixated image regions produced by ambient (exploratory) or focal processing modes. In the study of Velichkovsky et al (2005), and earlier in that of Velichkovsky (2002), it is proposed that longer saccades may stem from dorsal stream processes, which extend over the entire visual field, but are concerned with spatial and motor actuation (where/how) processes, whereas ventral stream processes operate over a restricted locale surrounding the high-spatial-resolution fovea, elicit shorter saccades, and are better supported by visual memory for appearance and semantic meaning. This study goes some way to supporting that position, also finding a correlation between saccade length and recognition rate, but additionally provides complementary evidence in that fixated image regions that were not subsequently recognised were not only, on average, produced by longer saccades but also typically exhibited significantly attenuated visual saliency.

Acknowledgments. This research was funded by NSF grants ECS-0225451 and ITR-0427372. We would like to thank Boris Velichkovsky and an anonymous reviewer for their very helpful comments.

References

Alvarez G A, Cavanagh P, 2004 "The capacity of visual short-term memory is set both by visual information load and by number of objects" *Psychological Science* **15** 106 – 111

Applied Science Laboratories, 1998 *Eye Tracking System Instruction Manual* (Version 1.2) (Bedford, MA: Applied Science Laboratories)

Baddeley R, 1996 "Searching for filters with 'interesting' output distributions: An uninteresting direction to explore?" *Network Computation in Neural Systems* **7** 409 – 421

Brainard D H, 1997 "The psychophysics toolbox" *Spatial Vision* **10** 433 – 436

Brewer W F, Treyans J C, 1981 "The role of schemata in memory for places" *Cognitive Psychology* **13** 207 – 230

Buswell G T, 1935 *How People Look at Pictures: A Study of the Psychology of the Perception of Art* (Chicago, IL: Chicago University Press)

Cowan N, 2001 "The magical number 4 in short-term memory: A reconsideration of mental storage capacity" *Behavioral & Brain Sciences* **24** 87 – 114

Crane H D, Steele C M, 1985 "Generation-V dual-Purkinje-image eyetracker" *Applied Optics* **24** 527 – 537

Creem S H, Proffitt D R, 1999 "Separate memories for visual guidance and explicit awareness", in *Stratification in Cognition and Consciousness"* Eds B H Challis, B M Velichkovsky (Amsterdam: John Benjamins) pp 73 – 96

Einhäuser W, König P, 2003 "Does luminance-contrast contribute to a saliency map for overt visual attention?" *European Journal of Neuroscience* **17** 1089 – 1097

Findlay J M, Gilchrist I D, 2003 *Active Vision: The Psychology of Looking and Seeing* (Oxford: Oxford University Press)

Gonzales R C, Woods R E, Eddins S L, 2004 *Digital Image Processing Using Matlab* Chapter 11 *Representation and Description* (Englewood Cliffs, NJ: Prentice Hall)

Haber R N, 1970 "How we remember what we see" *Scientific American* **222**(5) 104 – 112

Hateren J H, van, Schaaf A van der, 1998 "Independent component filters of natural images compared with simple cells in primary visual cortex" *Proceedings of the Royal Society of London, Section B* **265** 359 – 366

Henderson J M, 2003 "Human gaze control during real-world scene perception" *Trends in Cognitive Sciences* **7** 498 – 504

Henderson J M, Brockmole J R, Castelhano M S, Mack M, 2007 "Visual saliency does not account for eye movements during visual search in real world scenes", in *Eye Movements: A Window on Mind and Brain* Eds R van Gompel, M Fischer, W Murray, R Hill (Oxford: Elsevier) pp 537 – 562

Henderson J M, Hollingworth A, 1999 "The role of fixation position changes in detecting scene changes across saccades" *Psychological Science* **10** 438 – 443

Henderson J M, Hollingworth A, 2002 "Accurate visual memory for previously attended objects in natural scenes" *Journal of Experimental Psychology: Human Perception and Performance* **28** 113 – 136

Henderson J M, Weeks P A, Hollingworth A, 1999 "The effects of semantic consistency on eye movements during complex scene viewing" *Journal of Experimental Psychology: Human Perception and Performance* **25** 210 – 228

Henderson J M, Williams C C, Castelhano M S, Falk R J, 2003 "Eye movements and picture processing during recognition" *Perception & Psychophysics* **65** 725 – 734

Hollingworth A, 2006 "Scene and position specificity in visual memory for objects" *Journal of Experimental Psychology: Learning, Memory, and Cognition* **32** 58 – 69

Hollingworth A, 2007 "Object-position binding in visual memory for natural scenes and object arrays" *Journal of Experimental Psychology: Human Perception and Performance* **33** 31 – 47

Intraub H, Bender R S, Mangels J A, 1992 "Looking at pictures but remembering scenes" *Journal of Experimental Psychology: Learning, Memory, and Cognition* **18** 180 – 191

Intraub H, Richardson M, 1989 "Wide-angle memories of close-up scenes" *Journal of Experimental Psychology: Learning, Memory, and Cognition* **15** 179 – 187

Irwin D E, 1991 "Information integration across saccadic eye movements" *Cognitive Psychology* **23** 420 – 456

Irwin D E, 1992 "Memory for position and identity across eye movements" *Journal of Experimental Psychology: Learning, Memory, and Cognition* **18** 307 – 317

Irwin D, Andrews R V, 1996 "Integration and accumulation of information across saccadic eye movements", in *Attention and Memory XVI* Eds T Inui, J C McClelland (Cambridge, MA: MIT Press)

Irwin D, Zelinsky G J, 2002 "Eye movements and scene perception: memory for things observed" *Perception & Psychophysics* **64** 882 – 895

Itti L, Koch C, 2000 "A saliency-based search mechanism for overt and covert shifts of visual attention" *Vision Research* **40** 1489 – 1506

Itti L, Koch C, 2001 "Computational modelling of visual attention" *Nature Reviews Neuroscience* **2** 194 – 203

Jeannerod M, Rossetti Y, 1993 "Visuomotor coordination as a dissociable function: experimental and clinical evidence", in *Visual Perceptual Defects* Ed. C Kennard (London: Ballière Tindall) pp 439 – 460

Jiang Y, Olson I R, Chun M M, 2000 "Organization of visual short-term memory" *Journal of Experimental Psychology: Learning, Memory, and Cognition* **26** 683 – 702

Kahneman D, Treisman A, 1984 "Changing views of attention and automaticity", in *Varieties of Attention* Eds D R D Raja Parasuraman, R Davies (Orlando, FL: Academic Press) pp 29 – 61

Koch C, Ullman S, 1985 "Shifts in selective visual attention: towards the underlying neural circuitry" *Human Neurobiology* **4** 219 – 227

Li Z, 2001 "A saliency map in primary visual cortex" *Trends in Cognitive Sciences* **6** 9 – 16

Li Z, Snowden R, 2006 "A theory of a saliency map in primary visual cortex (VI) tested by psychophysics of colour-orientation interference in texture segregation" *Visual Cognition* **14** 911 – 933

Linde I van der, Rajashekar U, Bovik A C, Cormack L K, 2009 "DOVES: A database of visual eye movements" *Spatial Vision* **22** 161 – 177

Liu K, Jiang Y, 2005 "Visual working memory for briefly presented scenes" *Journal of Vision* **5** 650 – 658

Loftus G R, 1972 "Eye fixations and recognition memory for pictures" *Cognitive Psychology* **3** 525 – 551

Loftus G R, 1985 "Picture perception: Effects of luminance on available information and information-extraction rate" *Journal of Experimental Psychology: General* **114** 342 – 356

Luck S J, Vogel E K, 1997 "The capacity of visual working memory for features and conjunctions" *Nature* **390** 279 – 280

Mackworth N H, Morandi A J, 1967 "The gaze selects informative details without pictures" *Perception & Psychophysics* **2** 547 – 552

Maljkovic V, Martini P, 2005 "Short-term memory for scenes with affective content" *Journal of Vision* **5** 215 – 229

Melcher D, 2006 "Accumulation and persistence of memory for natural scenes" *Journal of Vision* **6** 8 – 17

Milner A D, Goodale M A, 1995 *The Visual Brain in Action* (Oxford: Oxford University Press)

Olson I R, Marshuetz C, 2005 "Remembering what brings along where in visual working memory" *Perception & Psychophysics* **67** 185 – 194

Olsson H, Poom L, 2005 "Visual memory needs categories" *Proceedings of the National Academy of Sciences of the USA* **102** 8776 – 8780

Parkhurst D J, Law K, Niebur E, 2002 "Modeling the role of salience in the allocation of overt visual attention" *Vision Research* **42** 107 – 123

Parkhurst D J, Niebur E, 2003 "Scene content selected by active vision" *Spatial Vision* **16** 125 – 154

Parkhurst D J, Niebur E, 2004 "Texture contrast attracts overt visual attention in natural scenes" *European Journal of Neuroscience* **19** 783 – 789

Pelli D G, 1997 "The VideoToolbox software for visual psychophysics: transforming numbers into movies" *Spatial Vision* **10** 437 – 442

Phillips W A, 1974 "On the distinction between sensory storage and short-term visual memory" *Perception & Psychophysics* **16** 283 – 290

Post R B, Welch R B, Bridgeman B, 2003 "Perception and action: Two modes of processing visual information", in *The Influence of H. W. Leibowitz* Eds J Andre, D A Owens (Washington, DC: American Psychological Association) pp 143 – 154

Privitera C, 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, Cormack L K, 2007 "Foveated analysis of image features at fixations" *Vision Research* **47** 3160 – 3172

Reinagel P, Zador A M, 1999 "Natural scene statistics at the centre of gaze" *Network: Computation in Neural Systems* **10** 341 – 350

Shepard R N, 1967 "Recognition memory for words, sentences and pictures" *Journal of Verbal Learning and Verbal Behaviour* **6** 156 – 163

Simons D J, Levin D T, 1997 "Change blindness" *Trends in Cognitive Sciences* **1** 261 – 267

Simons D J, Rensink R A, 2005 "Change blindness: Past, present and future" *Trends in Cognitive Sciences* **9** 16 – 20

Sperling G, 1960 "The information available in brief visual presentations" *Psychological Monographs: General and Applied* **74** 1 – 30

Sperling G, 1986 "A signal-to-noise theory of the effects of luminance on picture memory: Comment on Loftus" *Journal of Experimental Psychology: General* **115** 189 – 192

Standing L, 1973 "Learning 10,000 pictures" *Quarterly Journal of Experimental Psychology* **25** 207 – 222

Tatler B W, 2007 "The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions" *Journal of Vision* **7** 1 – 17

Tatler B W, Baddeley R J, Gilchrist I D, 2005 "Visual correlates of fixation selection: effects of scale and time" *Vision Research 45* 643 – 659

Tatler B W, Baddeley R J, 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

Torralba A, 2003 "Modeling global scene factors in attention" *Journal of the Optical Society of America A* **20** 1407 – 1418

Torralba A, Oliva O, Castelhano M S, Henderson J M, 2006 "Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search" *Psychological Review* **113** 766 – 786

Treue S, 2003 "Visual attention: the where, what, how and why of saliency" *Current Opinions in Neurobiology* **13** 428 – 432

Underwood G, Foulsham T, 2006 "Visual saliency and semantic incongruency influence eye movements when inspecting pictures" *Quarterly Journal of Experimental Psychology* **59** 1931 – 1949

Underwood G, Foulsham T, Loon E van, Humphreys L, Bloyce J, 2006 "Eye movements during scene inspection: A test of the saliency map hypothesis" *European Journal of Cognitive Psychology* **18** 321 – 342

Unema P, Pannasch S, Joos M, Velichkovsky B M, 2005 "Time-course of information processing during scene perception: The relationship between saccade amplitude and fixation duration" *Visual Cognition* **12** 473 – 494

Ungerleider L G, Mishkin M, 1982 "The two cortical visual systems", in *Analysis of Visual Behavior* Eds D J Ingle, M A Goodale, R J W Mansfield (Cambridge, MA: MIT Press) pp 549 – 586

vanRullen R, 2005 "Visual saliency and spike timing in the ventral visual pathway" *Journal of Physiology (Paris)* **97** 365 – 377

vanRullen R, Koch C, 2003 "Competition and selection during visual processing of natural scenes and objects" *Journal of Vision* **3** 75 – 85

Velichkovsky B M, 2002 "Heterarchy of cognition: The depths and the highs of a framework for memory research" *Memory* **10** 405 – 419

Velichkovsky B M, Joos M, Helmert J R, Pannasch S, 2005 "Two visual systems and their eye movements: Evidence from static and dynamic scene perception", in *Proceedings of the XXVII Conference of the Cognitive Science Society* Eds B G Bara, L Barsalou, M Bucciarelli (Mahwah, NJ: Lawrence Erlbaum Associates) pp 2283 – 2288

Wilcoxon F, 1945 "Individual comparisons by ranking methods" *Biometrics Bulletin* **1** 80 – 83

Wolfe J, 1998 "What do you know about what you saw?" *Current Biology* **8** 303 – 304

Yarbus A L, 1967 *Eye Movements and Vision* (New York: Plenum Press)

Zelinsky G J, Loschky L C, 2005 "Eye movements serialize memory for objects in scenes" *Perception & Psychophysics* **67** 676 – 690

ISSN 0301-0066 (print) ISSN 1468-4233 (electronic)

PERCEPTION

www.perceptionweb.com