Cue dynamics underlying rapid detection of animals in natural scenes

James H. Elder

Centre for Vision Research, York University, Toronto, Ontario, Canada

Ljiljana Velisavljević

Centre for Vision Research, York University, Toronto, Ontario, Canada

Humans are known to be good at rapidly detecting animals in natural scenes. Evoked potential studies indicate that the corresponding neural signals can emerge in the brain within 150 msec of stimulus onset (S. Thorpe, D. Fize, & C. Marlot, 1996) and eye movements toward animal targets can be initiated in roughly the same timeframe (H. Kirchner & S. J. Thorpe, 2006). Given the speed of this discrimination, it has been suggested that the underlying visual mechanisms must be relatively simple and feedforward, but in fact little is known about these mechanisms. A key step is to understand the visual cues upon which these mechanisms rely. Here we investigate the role and dynamics of four potential cues: two-dimensional boundary shape, texture, luminance, and color. Results suggest that the fastest mechanisms underlying animal detection in natural scenes use shape as a principal discriminative cue, while somewhat slower mechanisms integrate these rapidly computed shape cues with image texture cues. Consistent with prior studies, we find little role for luminance and color cues throughout the time course of visual processing, even though information relevant to the task is available in these signals.

Keywords: object recognition, natural scenes, contour, shape, texture, color

Citation: Elder, J. H., & Velisavljević, L. (2009). Cue dynamics underlying rapid detection of animals in natural scenes. *Journal of Vision, 9*(7):7, 1–20, http://journalofvision.org/9/7/7/, doi:10.1167/9.7.7.

Introduction

Given the complexity of natural scenes, humans are surprisingly quick at detecting the presence of specified types of object (e.g., animals, faces, cars). Constraints on visual processing time come from a combination of ERP and behavioral studies. In an early ERP study of brain activity during visual animal detection, Thorpe, Fize, and Marlot (1996) reported a stimulus-dependent differential signal as early as 150 msec following stimulus onset, suggesting very early availability of neural representations upon which discrimination could be based. By comparing ERP response to the same stimuli under different task conditions, VanRullen and Thorpe (2001b) demonstrated statistically significant *task-related* differential signals as early as 156 msec after stimulus onset, for both animal and vehicle targets. By changing the task from trial to trial, Johnson and Olshausen (2003) subsequently showed that *recognition-related* differentials appear as early as 152 msec after stimulus onset and found that the onset of the differential correlated with behavioral reaction time. Together, these ERP studies suggest that rapid object detection can be based upon neural signals emerging as soon as 150 msec after stimulus onset, although more difficult discriminations seem to involve longer processing times.

Behavioral studies have also put strong constraints on the speed of object detection in natural scenes. Employing a two-alternative forced-choice paradigm, Kirchner and Thorpe (2006) presented participants with simultaneous pairs of animal and non-animal images, randomly assigned to left and right hemifields. The participants' task was to rapidly direct gaze to the animal image. These experiments revealed mean saccadic reaction times of 228 msec, and a minimum saccadic latency yielding above-chance performance averaging 150 msec, but as low as 120 msec for some participants. Accounting for the motor component of this delay, Kirchner and Thorpe argued that for the shortest responses, visual processing must be limited to roughly 95–100 msec.

Based on these findings, it has been argued that visual processing underlying very rapid animal detection in natural scenes must be based upon feedforward (i.e., non-iterative, non-recurrent) computations (Kirchner & Thorpe, 2006; Thorpe et al., 1996). While it has been shown that very simple low-level image differences between animal and non-animal images (e.g., pixel intensity statistics, spectral density slopes) are not the basis for rapid animal detection (Kirchner & Thorpe, 2006), otherwise the neural basis for rapid object detection remains largely unknown. A key step toward understanding these mechanisms is to understand the visual cues on which they rely. This is the goal of the present paper, for the specific task of animal detection in natural scenes. We focus here on four potential cues: luminance, color, texture, and shape. First, we review

doi: 10.1167/9.7.7 Received September 15, 2008; published July 10, 2009 ISSN 1534-7362

previous efforts to understand the role played by these four cues in object and natural scene recognition.

Object recognition

There has been considerable debate regarding the role of shape (Biederman, 1987; Biederman & Ju, 1988) and surface properties (Bruner, 1957) in object recognition. For isolated objects, Biederman and Ju (1988) found no difference in reaction time latency or accuracy for naming objects in color photographs or line drawings and argued that surface characteristics (e.g., color and texture) play a secondary role in object recognition when bounding contours are readily extracted. Roughly contemporary studies using a different task showed no improvement in object recognition performance with color images over monochrome images (Ostergaard & Davidoff, 1985) or line drawings (Davidoff & Ostergaard, 1988), supporting Biederman's (Biederman, 1987; Biederman & Ju, 1988) assertions. However, other studies have contradicted these results, showing superior performance for color images compared with monochrome images (Wurm, Legge, Isenberg, & Luebker, 1993) or line drawings (Brodie, Wallace, & Sharrat, 1991; Humphrey, Goodale, Jakobson, & Servos, 1994; Price & Humphrey, 1989), suggesting that surface characteristics do play a role.

There are major methodological differences between these studies (Biederman, 1987; Brodie et al., 1991; Davidoff & Ostergaard, 1988; Humphrey et al., 1994; Ostergaard & Davidoff, 1985; Price & Humphrey, 1989), and the equivocal results may reflect stimulus differences. For example, Tanaka and Presnell (1999) have demonstrated a reliable color effect for objects that consistently appear in one color (e.g., lemons), whereas this effect is absent for objects that are typically seen in a variety of colors (e.g., cars). Thus, perhaps not surprisingly, color benefits categorization only if it is diagnostic of the object category (Humphrey et al., 1994; Tanaka & Presnell, 1999; Wurm et al., 1993). In summary, the literature suggests that although shape information may be a principal basis for object categorization, surface properties may be used if they afford direct information about object category.

However, it is not entirely clear that we can generalize from these prior experiments with isolated object stimuli to the detection of objects embedded in natural scenes. For one thing, it is quite possible that contextual cues of the scene may contribute to object detection. In a classic series of studies, Biederman (1972) and Biederman, Glass, and Stacey (1973) demonstrated that an object can be detected faster within a coherent scene context than within a scrambled scene context, suggesting that the global structure of the scene affects object processing. It has also been argued, based upon human imaging results, that scene and object information are processed by distinct visual mechanisms (Epstein, Graham, & Downing, 2003; Epstein & Kanwisher, 1998). Thus, to understand object processing in natural scenes, it may be helpful to understand the processing of natural scenes in general.

Scene recognition

There are at least two ways in which luminance, color, and texture information may play a role in natural scene recognition. They may serve as direct cues to the scene category, e.g., the red of a sunset might suggest a beach scene, the texture of grass might suggest a meadow. They may also support scene recognition indirectly by assisting segmentation, thereby facilitating computation of shape and scene layout information on which recognition may be directly based.

Neuropsychological studies (e.g., Steeves et al., 2004) provide some evidence that color can provide a direct cue to scene recognition. Further, based on a comparison between psychophysical results and a computational model of texture discrimination, Renninger and Malik (2004) have suggested that recognition of scenes presented very briefly (37-msec presentation time) may be based largely on texture recognition. However, they also found that the correlation between human performance and their texture model declined substantially for longer presentation times. These results suggest that multiple mechanisms may underlie scene recognition and that these mechanisms may have different dynamics.

As with isolated objects, there is evidence that color may only be useful for scene classification when it is diagnostic of scene category (Goffaux et al., 2005; Oliva & Schyns, 2000). For example, Oliva and Schyns (2000) showed that images for which color was diagnostic of scene category (e.g., canyons, forests) were best recognized in their normal colors, and that the luminance-only versions were better recognized than abnormal-color versions. In contrast, for scenes where color is not diagnostic (e.g., shopping area, bedroom), recognition performance was equivalent between normal-color, luminance-only, and falsely colored image conditions.

In summary, the experimental evidence suggests that color and texture information may contribute directly to scene recognition, but that other cues, such as shape and scene layout information, may also play a role, particularly for longer presentation times.

Animal detection

While little is known about the visual cues used by humans for detection of general objects in natural contexts, there is good empirical evidence that fast visual detection of animals in natural scenes does not depend upon color (Delorme, Richard, & Fabre-Thorpe, 2000; Fei-Fei, VanRullen, Koch, & Perona, 2005). Our goal here is to build on this work to understand the role of luminance, shape, and texture information as well. Testing the role of color at the same time permits a useful cross-validation of our method with previous studies.

General methods

Participants

We conducted three experiments, each with a different set of eight participants, for a total of 24 participants. All were naive to the goals of the experiment, had normal or corrected-to-normal vision, and received CAD $10/hour for their participation.

Apparatus

A 16-inch (32 × 24 cm) Sony Trinitron CRT (1024 × 768, 100 Hz) was used to display the stimuli. The experiments were programmed with Matlab 6.5 using the Psychophysics toolbox (Brainard, 1997; Pelli, 1997) for stimulus presentation and synchrony between the graphics card and the monitor.

Stimuli

All stimuli (fixation screen, test images, and mask images) were centered on a gray background. The fixation screen consisted of a black cross subtending 0.4° visual angle.

We made use of the publicly available component of the Berkeley Segmentation Dataset (BSD; Martin, Fowlkes, & Malik, 2004), which consists of 300 color Corel photographs that each contain at least one discernable object. The database also provides segmentations for each image produced by human participants (Figure 1). The segmentation instructions to the participants were to divide each image into pieces that represent distinguished things of equal importance. We selected for each image the human segmentation that produced the median number of segments. We then used these segmentations to create stimuli in which luminance, color, texture, and shape information could be manipulated independently in order to measure their respective roles in animal detection.

Specifically, from this database, we extracted a set of images containing one or more non-human animals (n = 88), a set containing no animals (n = 88), and a set containing humans (n = 45; Figure 2). Animal and non-animal scenes were used to create the test images, and people scenes were used as post-stimulus masking images. Test and mask images were 320 × 480 pixels, subtending 3.3 × 5.0 deg visual angle. Manipulations applied to the test images within a condition were also applied to the mask images for that condition. In addition, mask images were divided into 8 × 8 patches and block scrambled (Figure 3).

Since the size of the BSD is limited, participants saw each of the images more than once. To reduce systematic learning effects, we counterbalanced stimulus duration and the order of image conditions across participants.

Procedure

The two main independent variables of interest in our experiments were stimulus duration and stimulus manipulation (e.g., color, monochrome). Stimulus duration (30, 60, 90, or 120 msec) was blocked and counterbalanced across participants. Stimulus manipulation was blocked and nested in random order within each stimulus duration block. There were 40 trials per stimulus duration/stimulus manipulation condition. Within a condition, mask and test images were randomly sampled without replacement for each participant.

Participants were positioned with a head-chin rest 172 cm from the screen. Each trial consisted of a fixation stimulus (1 sec), a test image, a mask image (50 msec), and a query screen (until response; Figure 3). Participants pressed one of two keys on a computer keyboard to indicate whether the test image was an animal or a non-animal scene. Participants were given no instructions on how quickly to respond, and there was no time limit on their response. There was feedback: A tone sounded if the response was incorrect.

Experiment 1: Shape and surface cues

In Experiment 1, we employed eight manipulations of the BSD test images in an attempt to discriminate the respective roles of shape and surface (luminance, color, and texture) cues on rapid animal detection in natural scenes (Figure 4). Experiment 1 was thus an 8 (stimulus manipulation) × 4 (stimulus duration) within-subjects design.

Methods

Example stimuli are shown in Figure 4, and a list of their properties is provided in Table 1. Conditions a and e employ full color images. Conditions b and f employ monochrome images, derived from the original color images by averaging RGB channels. (In Experiment 2, we derive luminance using the standard Y'UV color space to assess sensitivity to the method employed to create the monochrome stimuli.) Conditions c and g employ what might be called "paint-by-number" images, produced by averaging RGB colors within segments. These images thus provide large-scale color and luminance cues as well as shape information. By not mixing color and luminance information across segment boundaries, we are, in a sense, privileging these colors and luminance cues, affording them greater opportunity to contribute to performance.

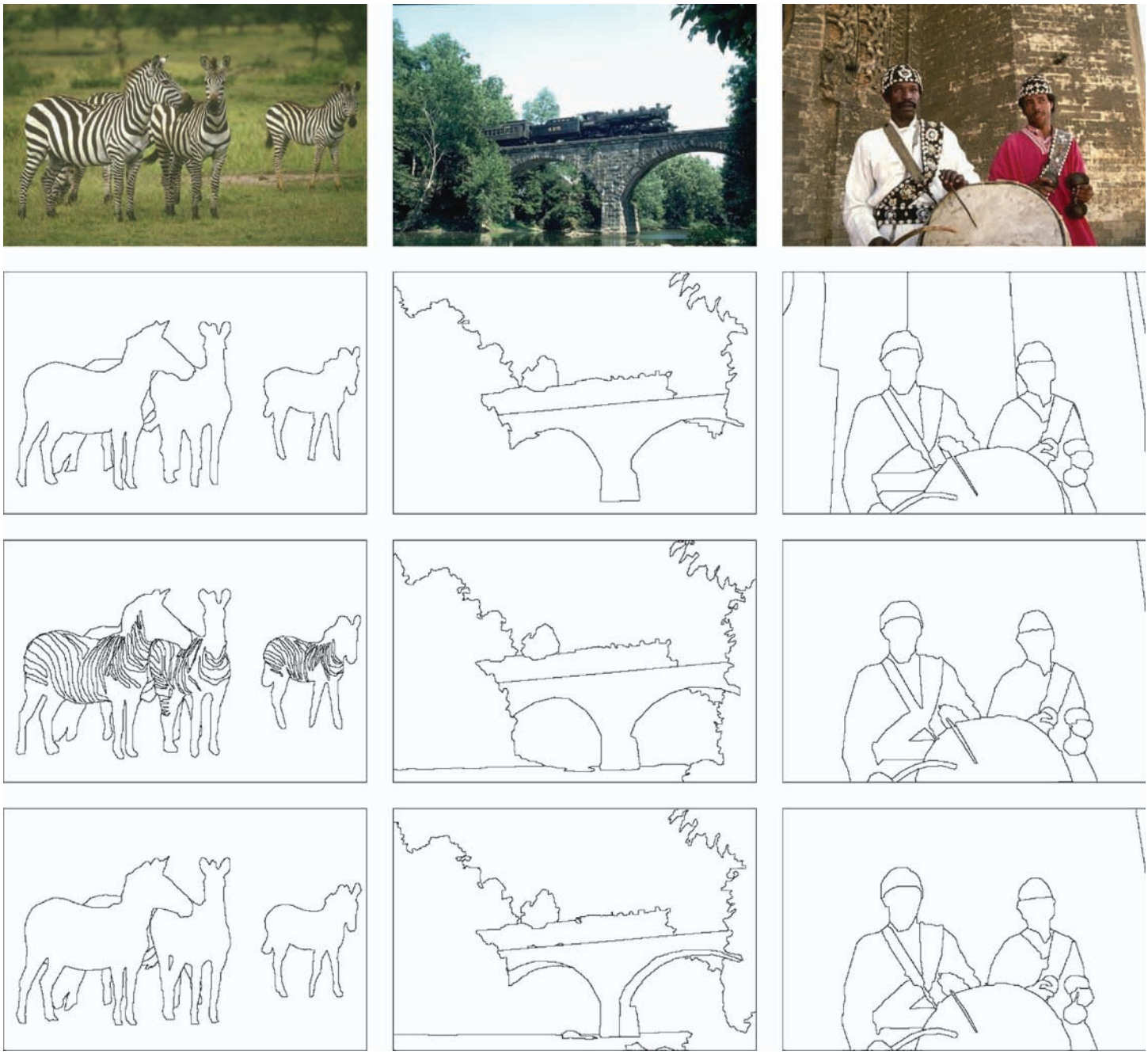

Figure 1. Sample photographs from the Berkeley Segmentation Dataset, with segmentations produced by three different participants.

In Condition d, only the segment outlines are displayed, so that shape is the only cue available for the task.

Condition h represents our attempt to produce stimuli that provide roughly localized luminance and color cues, but no shape or texture cues. Each of these stimuli consists of a Voronoi tessellation of the image based on the centers of mass of the image segments. Each cell of the tessellation represents the set of image points that are closest to the center of mass of one of the image segments. We painted the interior of each cell with the average RGB color of the corresponding segment. Thus, the Voronoi tessellation represents the colors and luminances of the original image in roughly the correct proportion and locations but provides essentially no shape information.

There is no universal definition for texture. In this paper, we operationally define texture as information available in the image beyond what is available in the shape of the segment boundaries and mean colors and luminances within these segments. It might be argued that some of the segment boundaries in the BSD afford texture information (for example, the second zebra segmentation in Figure 1). By employing only the segmentations with the median number of segments, we limit the amount of this texture information available in our non-texture conditions.

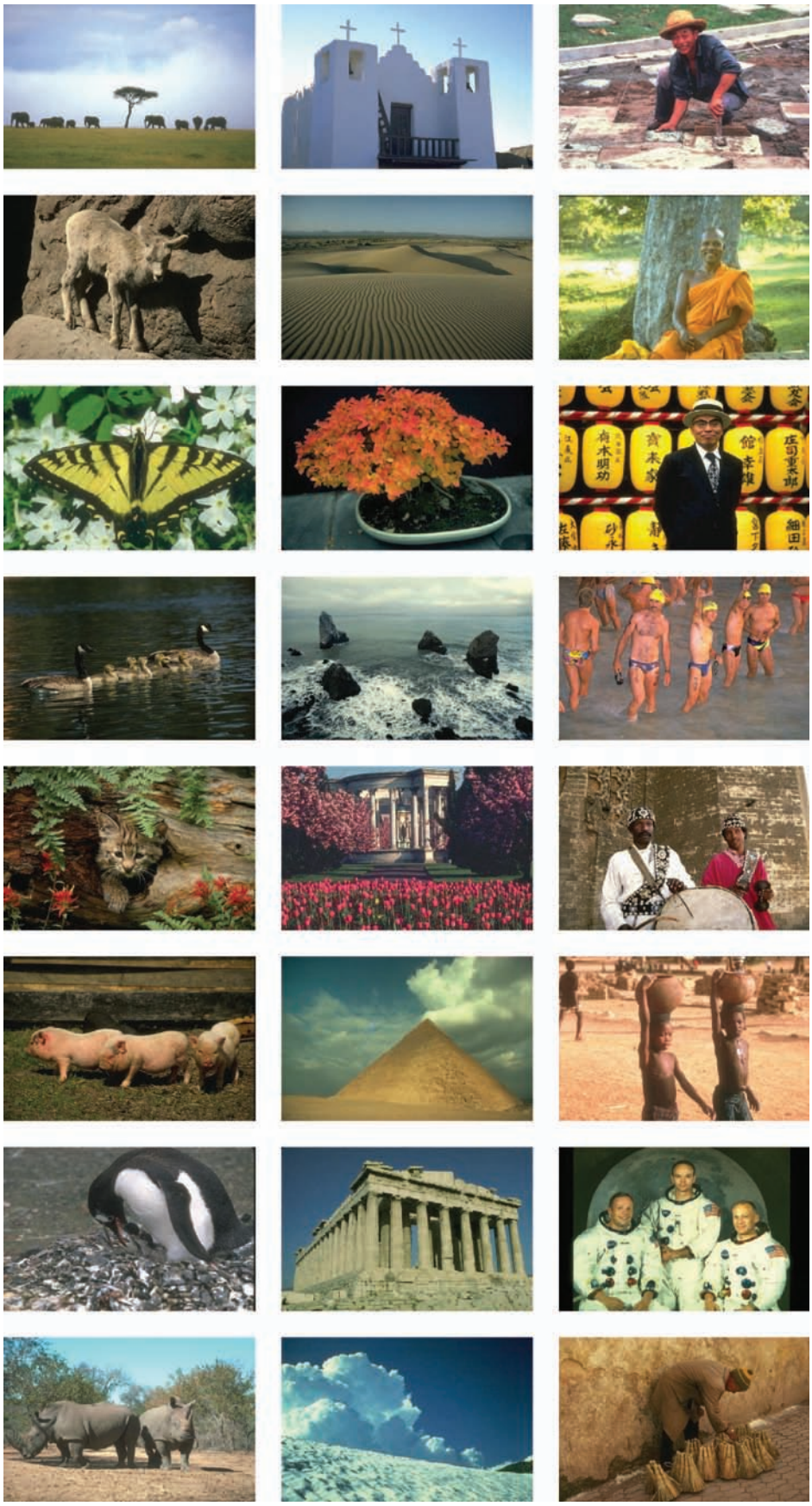

Figure 2. Examples of animal (left), non-animal (center), and people (right) images.

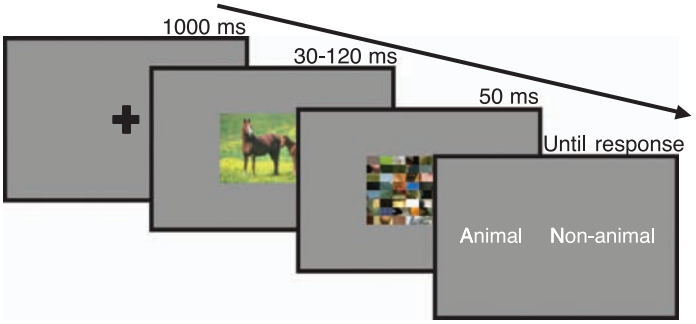

Figure 3. Trial sequence.

This operational definition of texture includes both luminance and color variations within segments. In the following, we make the assumption that the effective component of texture is carried by the luminance channel. In other words, we assume that isoluminant textures do not influence rapid animal detection. This assumption is based in part on evidence for poor discrimination of fine pattern information at isoluminance (Schiller, Logothetis, & Charles, 1991). Data from our first two experiments will allow us to assess the validity of this assumption.

In an effort to distinguish the effects of indirect cues that benefit recognition by facilitating segmentation, we overlaid on stimuli a–c 3-pixel-wide white contours depicting the segment boundaries (d). If color, luminance, or texture cues are contributing to recognition indirectly by facilitating segmentation, we would expect the effects of these cues to be greater for conditions e–g than for conditions a–c, where the segmentation is unambiguous.

For brevity, we refer to each stimulus using a string of letters representing the cues it affords: L = luminance, C = color, T = texture, S = shape. We use the letter O to indicate the emphasized segment outlines (Figure 4, Table 1).

Results

Figure 5 shows d' results for all eight stimulus manipulations and four stimulus durations t, averaged over the eight participants and presented in two separate panels for clarity. The data are generally well approximated by a thresholded linear model:

$$d' = r|\log_2(t/t_0)|^+, \quad (1)$$

where the notation $|\,|^+$ indicates half-wave rectification. (In other words, d' is constrained to be non-negative.) The model is governed by two parameters. The *cue delay* t_0 represents the minimum stimulus duration required for the participant to perform above chance, while the cue rate r reflects the rate at which sensitivity grows as stimulus duration is increased, specifically the amount by which d' increases when stimulus duration is doubled. (While using $\log_2 t$ provides an intuitive interpretation of d', to follow convention, we will plot d' as a function of $\log_{10} t$. The slopes differ by a factor of $\log_{10} 2$.)

Maximum likelihood estimates for the two parameters are shown in Figure 6. Confidence intervals for the parameters and all hypothesis tests are based on 1,000 bootstrapped samples over our eight participants, and we use these samples for all subsequent hypothesis tests. Cue rate was found to be significantly greater than zero for all conditions ($p < .001$) except the LC condition ($p = .08$).

Significantly positive cue rates for the SO condition indicate that boundary shape information alone is sufficient for rapid animal detection.

Comparing cue rates for LCTS versus LCS conditions indicates that texture speeds detection significantly ($p = .027$). A similar comparison for the outline stimuli (LCTSO vs. LCSO) shows a similar trend but does not reach significance ($p = .097$). This subdued effect of texture for the outline stimuli suggests that one role texture may be playing is in aiding segmentation. We will return to this question shortly.

The insignificant cue rate for the LC condition suggests that luminance and color do not play a large role in animal detection, but this does not mean they play no role. In particular, while performance is almost exactly at chance for short stimulus durations (30–60 msec), performance begins to rise above chance for longer stimulus durations (90–120 msec). Thus, the insignificant ($p = .08$) cue rate for the LC condition may in part be due to using two stimulus durations that are out of the range of processing for luminance and color cues in this task, thus effectively halving the data available for estimating the cue rate for the LC condition.

This interpretation is consistent with a comparison of the LCSO and SO conditions, which suggests a combined contribution of luminance and color cues late in processing (90–120 msec). A pairwise test between LCSO and SO conditions indicates cue rate is significantly greater for the LCSO condition ($p < .001$).

On the other hand, direct comparison of LCTSO versus LTSO and LCTS versus LTS conditions fails to show a significant effect of color by itself on cue rate ($p = .10$ and $p = .13$, respectively). This suggests that if color does play a role, it is minor role. It also lends credence to our assumption that the effective component of the texture signal is carried by the luminance channel (Methods section).

Taking these results together, it seems that color and/or luminance may have some influence late in the time course of stimulus processing, but this influence is fairly minor relative to the influence of shape and texture. These findings are consistent with prior reports suggesting little or no role for color in animal detection (Delorme et al., 2000; Fei-Fei et al., 2005).

There is also some variation in cue delay. First, note that since the cue rate for the LC condition is close to zero, the

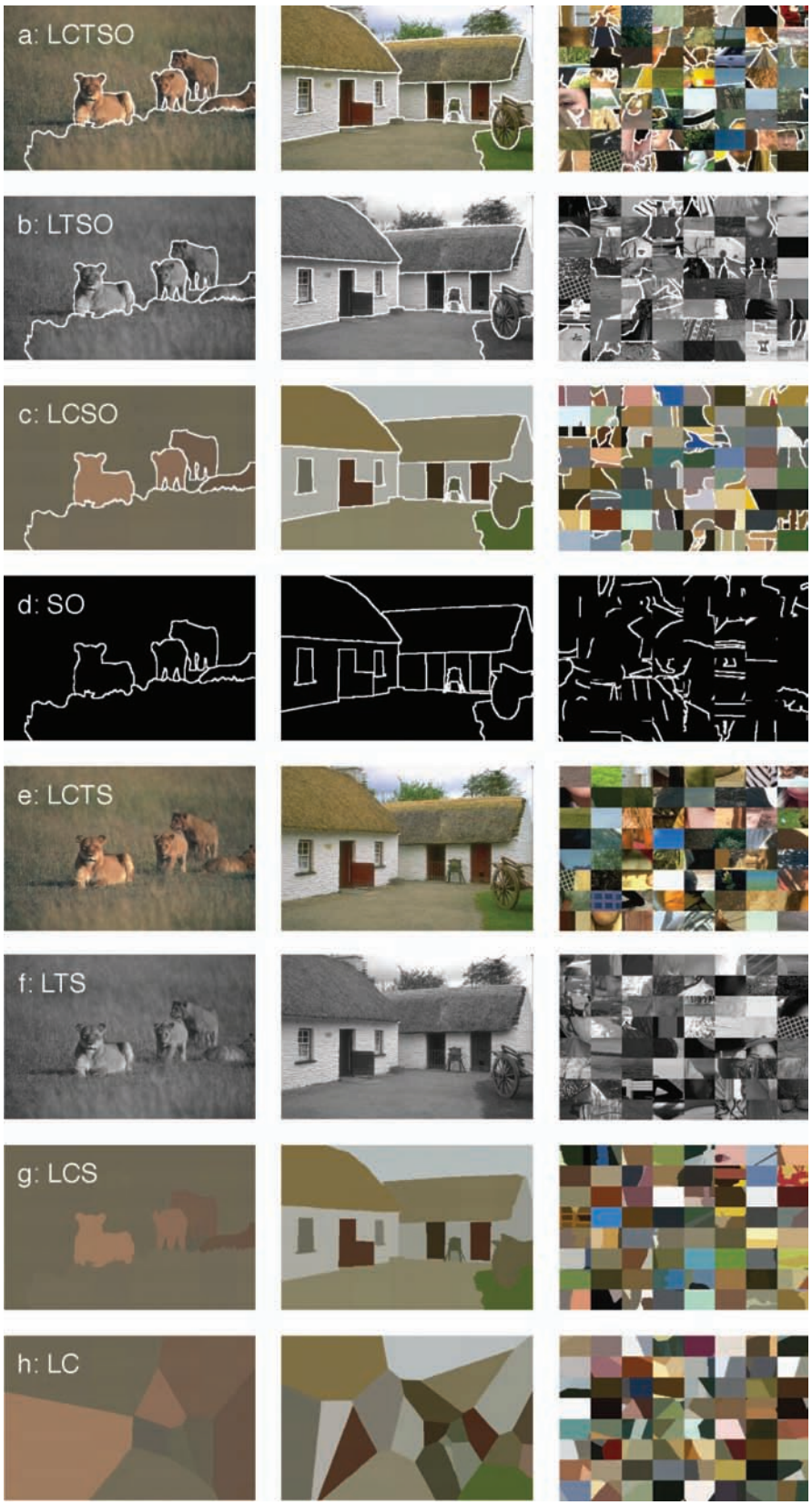

Figure 4. Example stimuli for animal (left) and non-animal (middle) scenes and the corresponding mask images (right) for each condition. The conditions are named according to the cues afforded by the image: L = luminance, C = color, T = texture, S = shape, O = outlined boundaries.

Condition	Luminance	Colour	Texture	Shape
LCTSO	✓	✓	✓	✓
LTSO	✓	✗	✓	✓
LCSO	✓	✓	✗	✓
SO	✗	✗	✗	✓
LCTS	✓	✓	✓	✓
LTS	✓	✗	✓	✓
LCS	✓	✓	✗	✓
LC	✓	✓	✗	✗

Table 1. Stimulus properties.

estimated cue delay is unreliable for this condition. It appears that cue delay may be shorter for the SO condition (contours only) than for other conditions; however, we postpone quantitative evaluation of the dynamics of individual cues to the Discussion section.

Finally, by comparing left and right panels of Figures 5 and 6, we observe that, qualitatively at least, the outline manipulation appears to interact with the other cues. The addition of the outline appears to improve detection rates when texture is absent (LCSO vs. LCS), presumably reflecting facilitated computation of shape information used for animal detection. However, the addition of the outline cue appears to *reduce* detection rates for images containing texture (LCTSO vs. LCTS and LTS vs. LTSO), perhaps due to partial masking of the useful texture information. We will consider this issue in more detail in the Discussion section.

Experiment 2: Color and texture

In Experiment 1 we found relatively minor effects of subtracting color from the stimuli (LCTSO vs. LTSO and LCTS vs. LTS), suggesting that color has little effect on animal detection. Although this is consistent with prior results on rapid animal detection (Delorme et al., 2000;

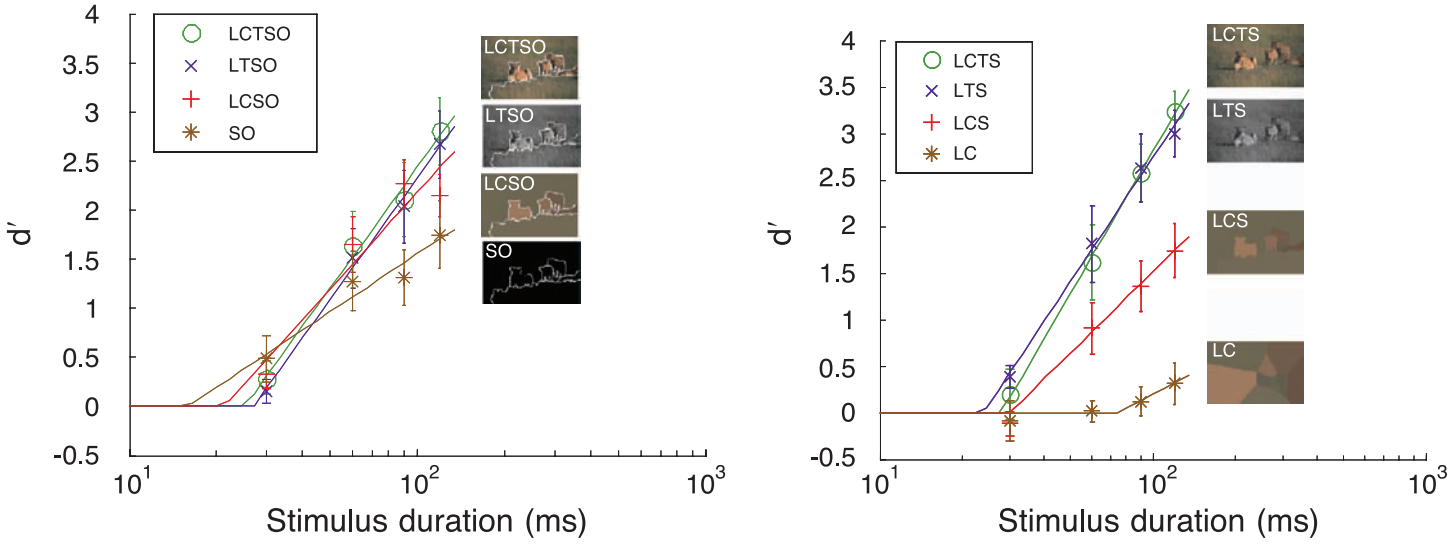

Figure 5. Recognition performance (Experiment 1). Lines represent maximum likelihood fit of rectified linear model. Error bars represent 1 standard error of the mean (*SEM*).

Fei-Fei et al., 2005), other work suggests that color may be a useful cue under particular circumstances (Oliva & Schyns, 2000; Roberts & Mazmanian, 1988; Tanaka & Presnell, 1999).

One possible concern with our experiment is a possible interaction between color and texture cues. For example, it is possible that color and texture cues are statistically correlated, so that the presence of texture masks the effects of color. A second concern with our first experiment is that the monochrome stimuli were derived by averaging RGB channels, which is only a rough approximation of how humans perceive luminance. Our second experiment therefore repeats three of the conditions from Experiment 1 but adds a fourth (LSO) that completes a crossed-design

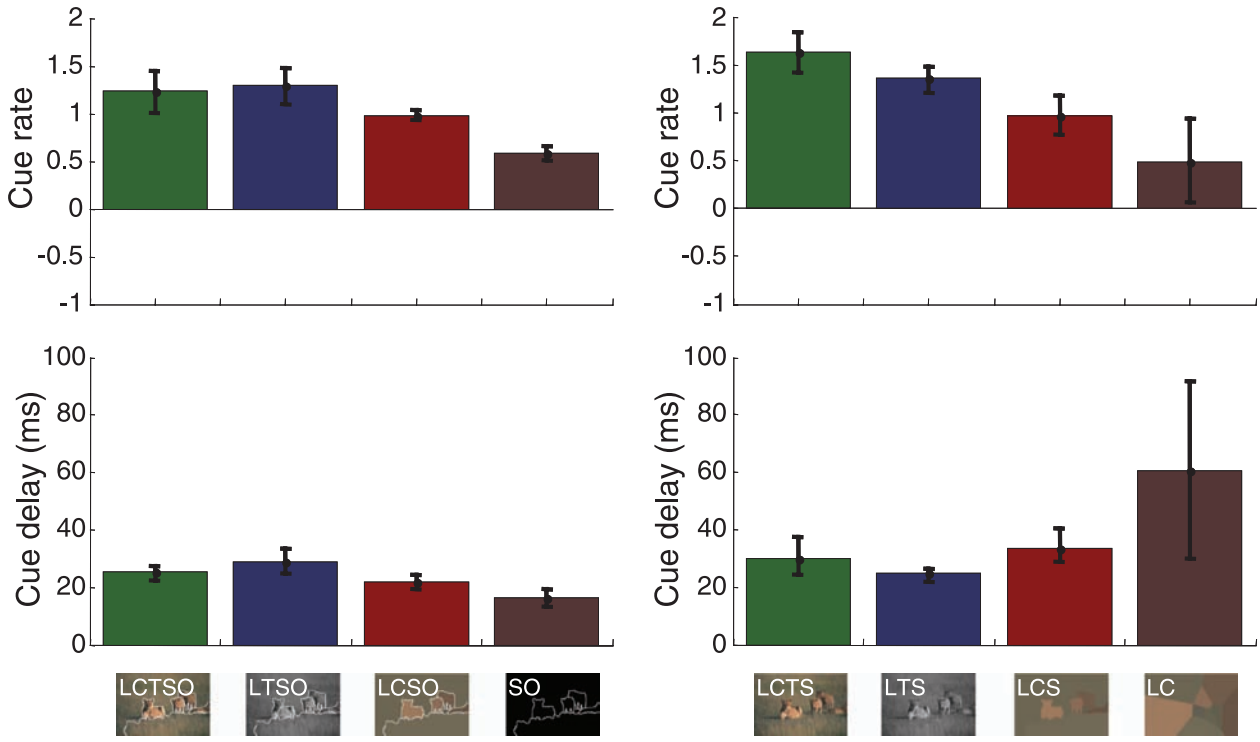

Figure 6. Cue rate (top) and cue delay (bottom) for each condition. Error bars represent the 68% confidence interval for the mean.

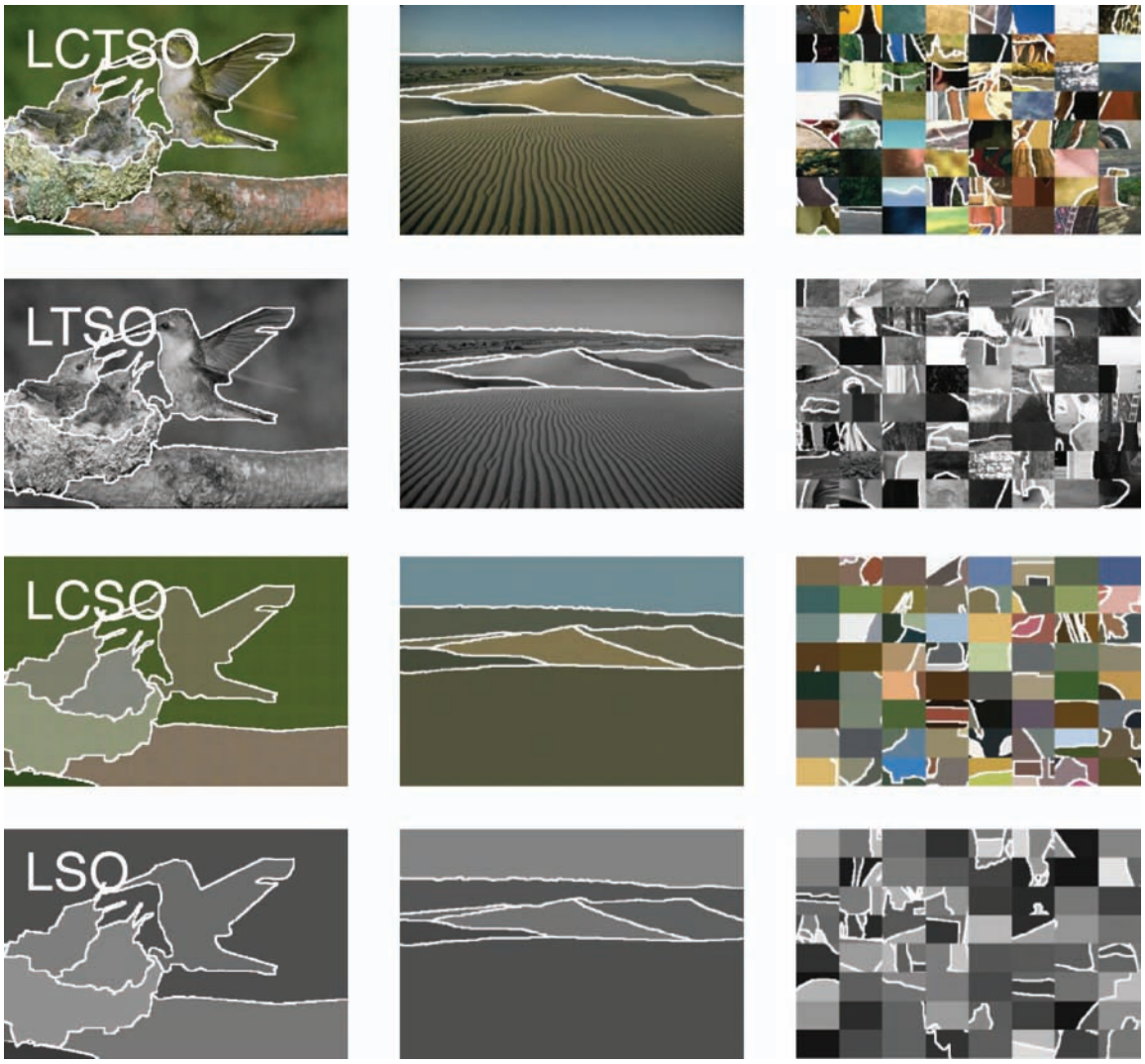

Figure 7. Example test images for animal scenes (left column) and non-animal scenes (middle column) and the corresponding mask images (right column) for each condition: (a) color, (b) gray, (c) paint-by-numbers color, and (d) paint-by-numbers gray conditions.

jointly testing the independent roles of color and texture in rapid animal detection and uses the standard Y'UV color space to derive monochrome stimuli. We use outlines in all conditions to make segmentation unambiguous, thus focussing the experiment on the direct role of color and texture in animal detection.

Methods

In all conditions, image segments were outlined with a white three-pixel contour (Figure 7). Monochrome images were derived using the MATLAB rgb2gray function, which is based upon a standard Y'UV color space representation of the original stimuli. Masks were randomly sampled from trial-to-trial for each participant without replacement. Thus, Experiment 2 was a 2 Color (Present, Absent) × 2 Texture (Present, Absent) design.

Results

The results are shown in Figure 8, and estimated cue delay and cue rate parameters are shown in Figure 9. There is clearly a main effect of texture on cue rate (p = .018). Since all stimuli included outlines highlighting segment boundaries, it appears likely that texture is serving as a direct cue rather than simply assisting segmentation. On the other hand, there is no main effect of color (p = .35) and no marginal effects of color (LCTSO vs. LTSO, p = .20, LCSO vs. LSO, p = .56). There are no significant differences in cue delay between the conditions ($p > .05$). Thus, we confirm the findings of Experiment 1: color does not appear to play a significant role in the rapid detection of animals in natural scenes, and the effective component of the texture signal is carried by the luminance channel (Methods section).

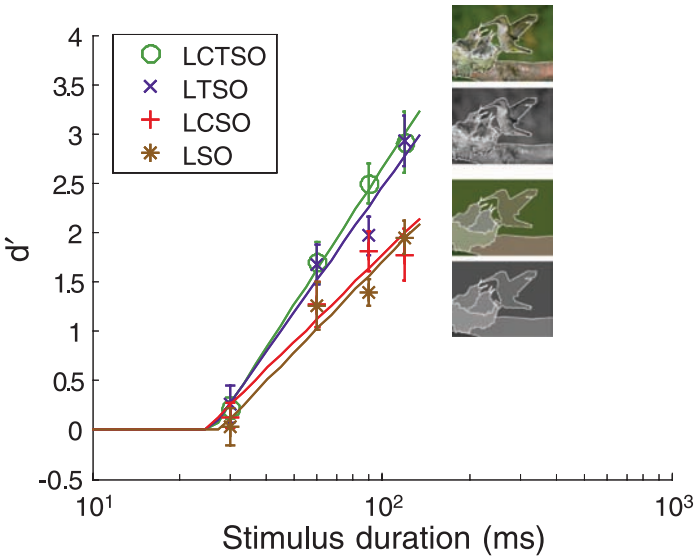

Figure 8. Experiment 2 results. Lines represent maximum likelihood fit of rectified linear model; error bars represent 1 *SEM*.

Repetition of Condition LTSO in both Experiments 1 and 2 allows assessment of sensitivity to the method for deriving monochrome images (averaging RGB intensity vs. the Y' channel of Y'UV color space). We find that results are quite independent of the method: neither the cue rate ($p = .47$) nor the cue delay ($p = .32$) are significantly different.

Experiment 3: Contours and edge maps

Our first experiment suggests that shape alone is a major cue for express detection of animals in natural scenes: To support rapid recognition, the visual system may thus rely upon a rapidly extracted contour representation of the scene.

It is not obvious, however, that the human visual system would be able to rapidly extract a contour representation that selectively demarcates the major salient structures of the image in the manner of the outline figures used in Experiment 1 (Figure 4). While these outlines were produced by human participants, these participants were not under any time restrictions and may have used substantial domain-knowledge, top-down cues, and color, texture, and luminance features to refine their segmentations. By comparison, machine-generated edge maps computed solely from low-level cues are typically cluttered and/or incomplete (Figure 10). These limitations of low-level edge maps are what makes the computer vision problems of contour completion and contour grouping so challenging (Elder, Krupnik, & Johnston, 2003; Parent & Zucker, 1989; Sha'ashua & Ullman, 1988; Williams & Thornber, 1999). Without complete salient contours, object recognition may be difficult. Indeed, Sanocki, Bowyer, Heath, and Sarkar (1998) reported that human object recognition performance dropped by a factor of more than two when edge map stimuli produced with the standard Canny (1986) algorithm were substituted for the original color photograph stimuli.

The question of whether bottom-up edge maps form a reliable basis for animal detection in natural scenes thus remains open. The purpose of our third experiment is thus to explicitly compare performance for our outline stimuli with performance for more realistic edge maps produced by a contemporary computer vision edge detection algorithm.

Methods

The multiscale edge detector we employ here (Elder & Zucker, 1998) is designed to detect all luminance transitions in the image not due to sensor noise. The chief parameter is an estimate of the standard deviation σ_n of this noise. We can use this parameter to generate edge maps with various levels of detail. Small values of σ_n lead to dense edge maps with copious detail, and large values

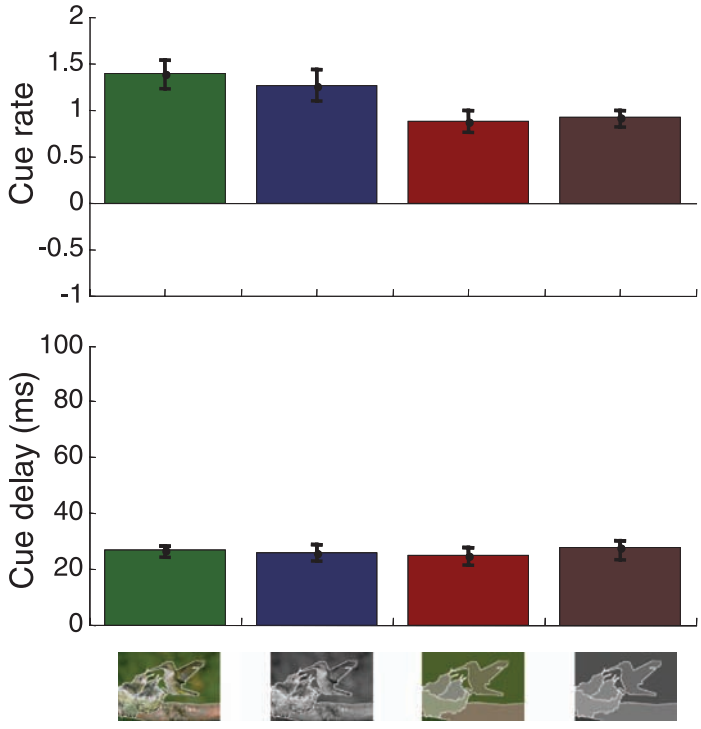

Figure 9. Cue rate (top) and cue delay (bottom) for each condition. Error bars represent 68% confidence intervals for the mean.

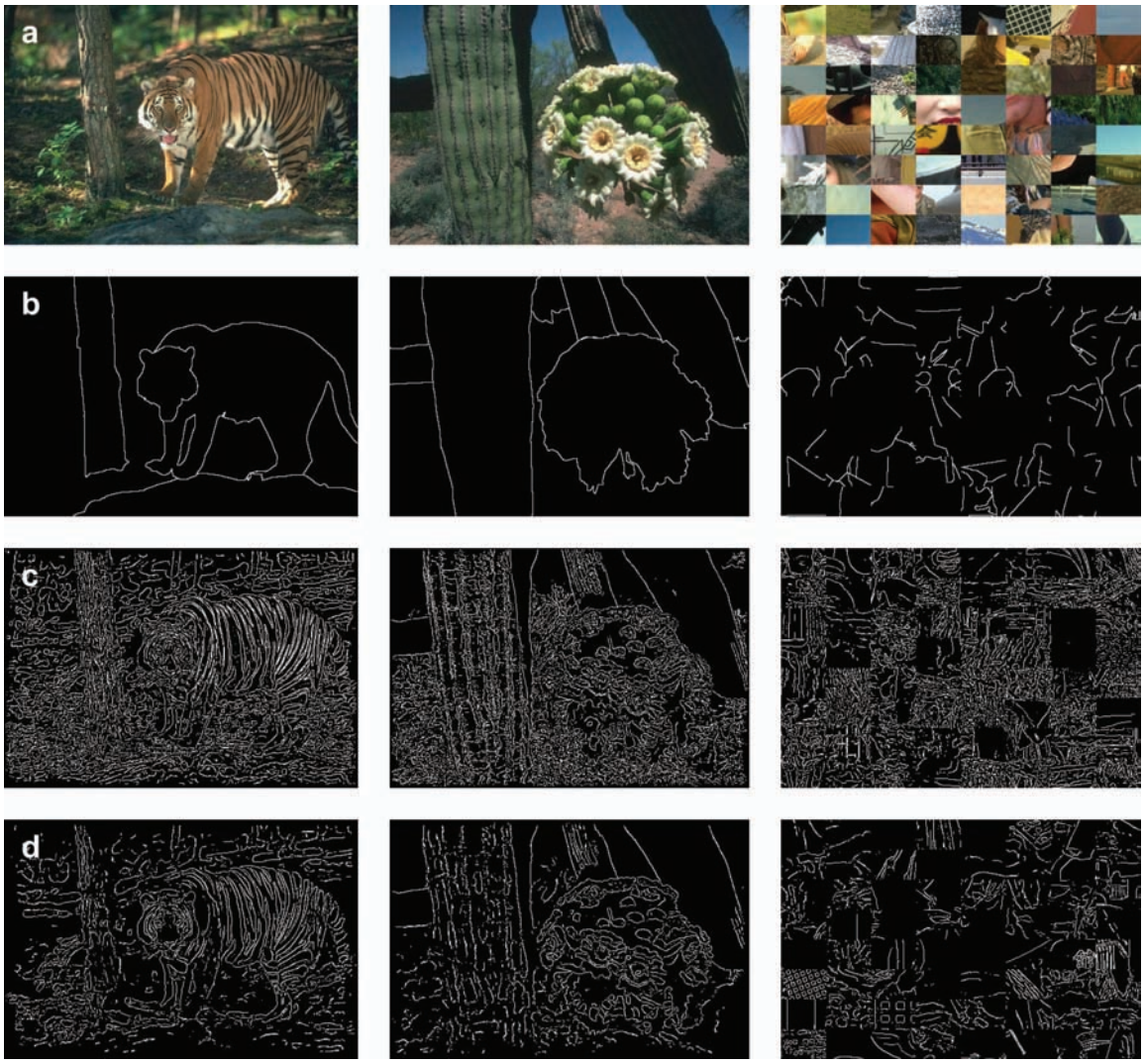

Figure 10. Example animal (left column), non-animal (middle column), and mask stimuli (right column) for each condition of Experiment 3: (a) color, (b) outline, (c) dense edge map, (d) sparse edge map.

of σ_n yield sparser edge maps in which low-contrast and/or highly blurred edges are suppressed.

For this experiment, we produced 1-pixel thick edge map stimuli at two levels of detail: dense (σ_n = 3 gray levels) and sparse (σ_n = 11 gray levels). For comparison, we also tested recognition for 1-pixel thick versions of the outline stimuli and the full color stimuli employed in Experiment 1. Mask images (Figure 10) matched the modality of the test images.

Results

Figure 11 shows the results of this experiment, and Figure 12 shows maximum likelihood estimates of cue rate and cue delay parameters. Cue rate is significantly greater than zero for all conditions ($p < .001$) and significantly greater for the complete image (LCTS) condition than for any of the contour conditions ($p < .001$), indicating that features other than shape (e.g., texture) certainly play a role.

Of greatest interest are the differences between the manual segmentation condition (SO) and the machine-generated edge map conditions (Dense and Sparse). Estimated cue rate for the SO condition was slightly greater than for the Dense and Sparse edge map conditions, but these differences were not significant ($p = .16$ and $p = .12$, respectively). Cue delay for all of the contour conditions (SO, Dense, Sparse) tended to be less than for the full image (LCTS): p = .001, .057, .034, respectively. The cue delays for the Dense and Sparse machine-generated edge maps were not significantly different from the cue delay for the manual segmentation (SO) condition: $p = .44$ and $p = .56$ respectively. Thus, we conclude that, while the clean outlines provided by the SO stimuli may lead to slightly better performance for longer

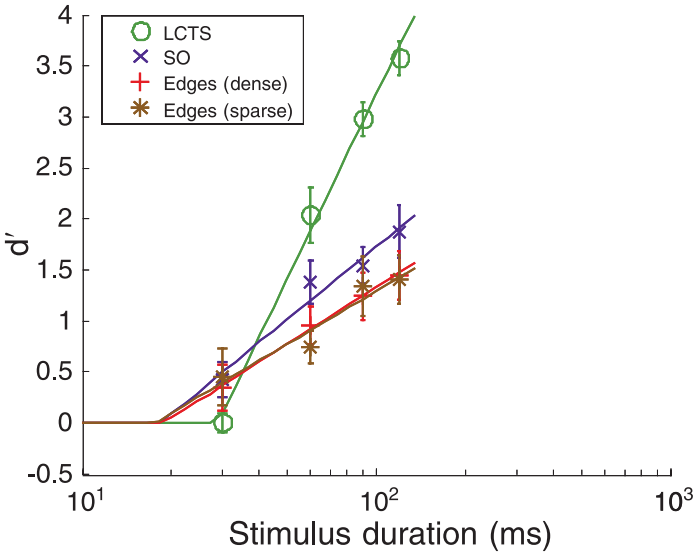

Figure 11. Experiment 3 results. Lines represent maximum likelihood fit of rectified linear model. Error bars represent 1 *SEM*.

stimulus durations, the great majority of the shape information provided is readily extracted from bottom-up edge maps and just as rapidly.

Estimating the time course of individual cues

A main goal of this work is to estimate how individual cues to animal detection in natural scenes are processed over time. Unfortunately, it is difficult to experimentally isolate each cue in natural image stimuli. Here we will use a simple regression model that will allow us to estimate the role of individual cues from experiments in which various combinations of these cues are present.

In all three of our experiments, we found that a simple thresholded log-linear model provides a good account of human performance as a function of time:

$$d' = |r\log_2(t/t_0)|^+. \quad (2)$$

Here we propose a simple superposition assumption to account for performance in terms of individual cue processes:

$$d' = |\sum_i r_i \log_2(t/t_i)|^+, \quad (3)$$

where r_i and t_i are the rate and delay for cue i.

When stimulus duration exceeds cue delay, this assumption leads to linear summation over cues. However, cues with longer delays will tend to have a subtractive effect on performance for shorter stimulus durations. Thus, the model is capable of accounting for delay due to competition between cues for shared cortical resources.

Fitting this mathematical model to the data from our three experiments could in principal allow us to estimate the cue rate and delay for each cue. We observed in Experiment 1, however, that the outline cue (highlighting of major boundaries) appears, at least qualitatively, to interact with other cues in a non-additive way. For example, highlighting boundaries appeared to facilitate contour processing while inhibiting texture processing. For this reason, we estimated cue parameters separately for conditions that include the outline cue and those that do not. We considered all conditions of Experiments 1–3 except for the sparse and dense edge conditions of Experiment 3. Some stimulus conditions were repeated in two separate experiments: In these cases, data from both experiments were used. We computed a least-squares fit in each case using a gradient-descent method, with multiple repetitions from random initial parameters to avoid local minima.

Figure 13 shows the fit to the data and estimated time course for each individual cue, for conditions including the segment outlines. Figure 14 shows the same for conditions without the outline cue. In both cases, the mathematical model provides a reasonable fit to the data: While there may be interactions between cues not accounted for by our simple mathematical model, the data do not reveal them.

From these two figures, it can be seen that while both shape and texture profoundly influence performance, color

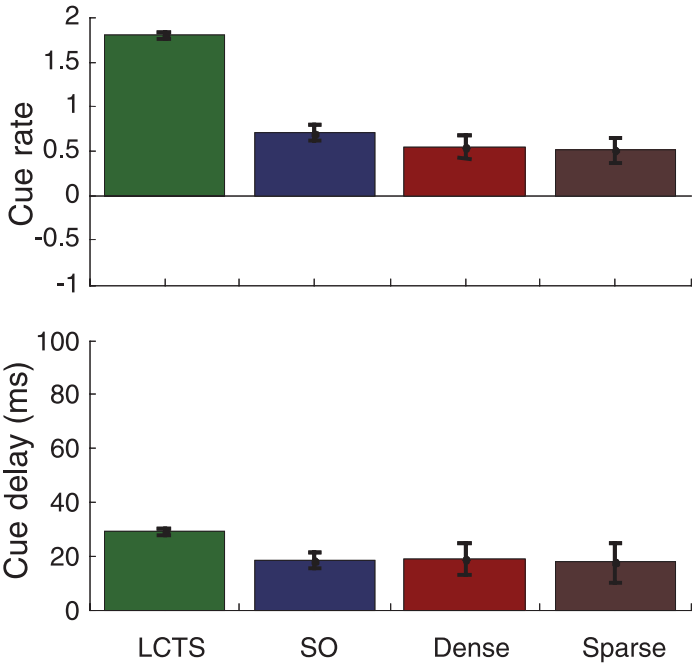

Figure 12. Cue rate (top) and cue delay (bottom) for each condition. Error bars represent 68% confidence intervals for the mean.

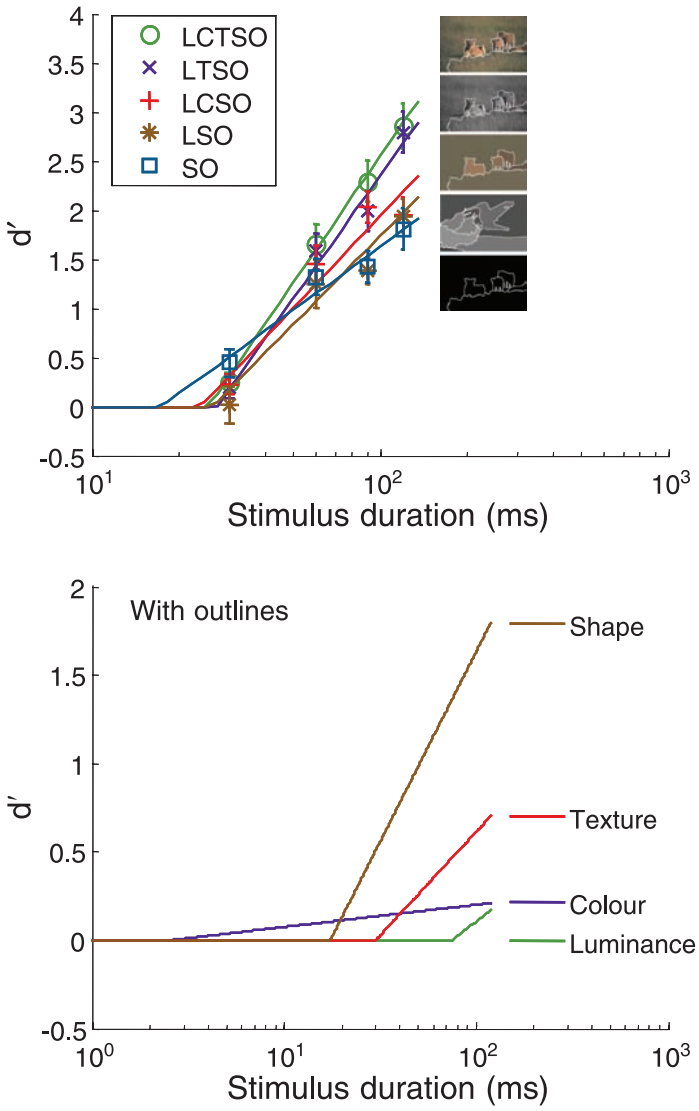

Figure 13. Top: Least-squares fit of multicue model to data for outline stimuli. Error bars indicate *SEM*. Bottom: estimated time course of individual cues.

and luminance have little effect, as measured by d' for this task. Table 2 quantifies these observations: Cue rate and cue delay are reported for each cue, and the significance of the cue rate parameter is computed based upon 1,000 bootstrapped samples over participants. While cue rates for both shape and texture are significantly positive ($p < .05$) for both outline and no-outline conditions, cue rates for luminance and color are not significantly different from 0 ($p > .05$) for either outline or no-outline conditions. Due to the relatively minor role of color and luminance in detection and the non-linear relationship between the cue rate and the delay parameters and d' for the task, the estimates of cue rate and cue delay for the color and luminance cues are not reliable.

The dominance of shape cues over texture cues for outline but not non-outline stimuli suggests that the outlining may ease extraction of the shape cues, reducing the utility of texture cues for segmentation and possibly masking texture cues useful for direct discrimination. Thus, we conclude that the key cues to animal detection in natural scenes are shape and texture, and that texture may play a dual role: facilitating segmentation but also providing direct cues for animal detection.

Discussion

Dynamics

We found a strong effect of stimulus duration in all of our experiments. For example, performance for complete

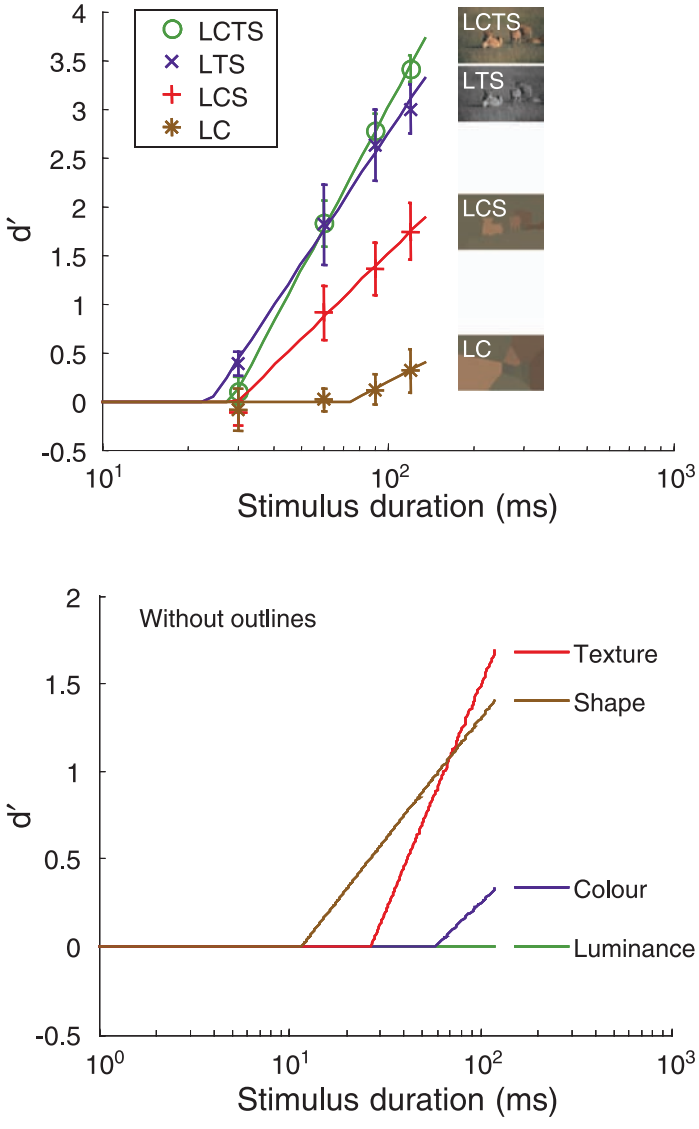

Figure 14. Top: Least-squares fit of multicue model to data for stimuli without outlines. Error bars indicate *SEM*. Bottom: estimated time course of individual cues.

	Cue rate	Significance (p)	Cue delay (msec)
With outlines			
Luminance	0.25*	.06	75*
Color	0.04*	.12	2.5*
Texture	0.35	<.001	30
Shape	0.65	<.001	17
Without outlines			
Luminance	0*	.47	—
Color	0.32*	.10	58*
Texture	0.79	.008	27
Shape	0.42	.035	12

Table 2. Estimated cue parameters. *Note*: Since the effects of color and luminance were found to not be statistically significant, the corresponding estimates of cue rate and cue delay are not reliable.

color natural image stimuli (LCTS) improved from roughly 55% correct at 30 msec duration to 94% correct at 120 msec duration. Comparison with earlier results highlights the importance of the post-stimulus mask: Using unmasked test images and a stimulus duration of 20 msec, Thorpe et al. (1996) and VanRullen and Thorpe (2001a) measured performance at over 90%. This comparison suggests that due to visual persistence and/or iconic memory (Averbach & Coriell, 1961; Neisser, 1967; Sperling, 1960), the effective duration of a briefly presented, unmasked natural image is at least 100 msec.

The dynamics revealed in our experiments can more easily be compared with previous results in natural scene categorization. Using a post-stimulus mask, Renninger and Malik (2004) found an increase in scene categorization performance from roughly 70% at 29 msec stimulus duration to 95% at 69-msec stimulus duration. This comparison suggests that human participants are faster/better at rapid scene categorization than animal detection.

It has been estimated (Serre, Oliva, & Poggio, 2007) that significant recurrent processing for masked stimulus detection should begin to play a role for stimulus durations longer than 40–60 msec. Thus, while cue delay estimates in the range of 15 msec for shape and 30 msec for texture (Table 2) suggest that purely feedforward mechanisms do provide discriminative power, it is quite likely that the continued increase in d' for longer stimulus durations derives in part from recurrent computations.

In our experiments, sensitivity (d') for animal detection was measured as a function of stimulus duration. Although participants were given no instructions regarding how quickly to respond and were given an unlimited amount of time for their response, we did record reaction time on each trial. Mean reaction times over conditions ranged from 678 to 861 msec. For each of the three experiments, we ran a three-way ANOVA with reaction time as the dependent variable, participant as a random factor and condition and stimulus duration as fixed factors. In each case, we found a significant main effect of participant ($p < .05$) but not condition or stimulus duration ($p > .05$).

Additive model

We assumed a simple superposition model to account for how multiple cues combine to determine performance (Discussion section). The model is additive: d' for the combined-cue stimulus is predicted to be equal to the sum of the d's for the constituent cues. However, since cue delay is non-linearly related to d', the model can potentially account for either increases or decreases in delay when cues are combined.

The model assumes that cues are independent. While it is quite likely that interactions exist, the present data provide no justification for elaborating the model further: The fit of the model to the data (Figures 13 and 14) is excellent.

Direct and indirect contributions of cues

Luminance, color, and texture cues can potentially contribute to animal detection in at least two ways. First, they can contribute to the segmentation of the image, thus yielding the shape information on which discrimination is based. (We take it as given that the luminance signal is important for segmentation, but the role of texture and color is not assumed.) Second, they may serve as direct cues to animal detection. For example, striped fur might suggest a tiger.

In Experiment 1, we attempted to distinguish these two issues for color and texture by employing both outline (O) and non-outline stimuli. In the outline stimuli, the major segment boundaries are emphasized by highly visible white contours that essentially solve the segmentation problem. If the main role of a particular cue is in aiding the segmentation process, we should find that the effect of that cue is diminished in the outline conditions, relative to the non-outline conditions.

In the following, we consider the direct roles played by shape, texture, color, and luminance cues as well as possible indirect roles played by texture and color in aiding segmentation.

Shape

The isolation of shape cues relies upon the human segmentation boundaries provided by the BSD. We are thus assuming that the shapes demarcated by the BSD participants are those that would prove useful for participants in our task. By selecting the BSD segmentation with the median number of segments, we hope at least to have a reasonably representative example of how these images are perceived given unlimited viewing time.

The relatively strong performance of our observers when only these outlines were available indicates that these segmentations are also highly useful for rapid animal detection. In fact, cue delays were found to be shortest for the shape cue (12–17 msec). Does this tell us how shape cues are used for unaltered natural image stimuli? The difficulty is that in isolating the shape cue, we relied upon hand segmentations created by participants given an unlimited amount of time. It is not clear that these shape cues could be extracted from natural images by participants in our task given the very brief stimulus exposures.

Experiment 3 addressed this question by employing edge map stimuli produced by a simple algorithm based upon visual filters that could plausibly be implemented by feedforward visual mechanisms in early visual cortex. Performance with these bottom-up edge stimuli was very similar to performance for the BSD contour stimuli. Most importantly, the cue delay was virtually identical for the outline and edge map conditions. This shows that this it is feasible for the shape cues on which rapid animal detection is based to be extracted efficiently from natural images by simple visual mechanisms in early visual cortex and lends further credence to shape-based models of object detection and recognition (Biederman & Ju, 1988; Ullman, 1989).

Texture

We employed an operational definition of texture as all image variation within segments of the BSD. While there is no perfect definition of texture, this definition at least has the advantage of being simple and unambiguous.

For the purposes of this study, we assumed that the effective component of the texture cue is carried by the luminance channel. This assumption is supported by the results of Experiment 2: Performance is almost identical for LCTSO and LTSO conditions and for LCTS and LTS conditions.

Texture as defined was found to be an important factor in the detection of animals in natural scenes. However, we estimate that computation of texture information requires longer stimulus exposure (27–30 msec) than for shape.

In Experiment 1, we found that performance was better for textured images without outlines than with and the estimated cue rate for texture more than doubled (from .35 to .79) when the outline cues were removed from the imagery. These findings could reflect partial masking of the texture information by the outlines and a role for the texture cues in segmentation. However, the results of Experiment 2 show that texture improves performance even when outlines are present, suggesting that texture also provides a direct cue to animal detection, beyond any role played in segmentation.

It has been suggested (Renninger & Malik, 2004) that very rapid scene categorization may be based largely on texture cues. Our results suggest that animal detection is somewhat different: Shape cues are computed most rapidly, and texture cues follow. Our results can be compared with the findings of Vogel (1999a, 1999b), who found that texture alone could not account for performance on fish/non-fish and tree/non-tree categorization in rhesus monkey.

Color and luminance

Our first two experiments confirm prior studies (Delorme et al., 2000; Fei-Fei et al., 2005) showing little effect of color on rapid animal detection in natural scenes. One possible reason for this null effect is that color is not generally a discriminative feature for animals in natural scenes. To explore this possibility, we examined the color statistics for the 88 animal and 88 non-animal LCTS test images used in Experiment 1. We experimented with several color spaces and found that the L*a*b* color space provided the most discriminative color information (see below). In the L*a*b* color space, the L* channel carries the luminance signal, while the a* and b* channels carry green/magenta and blue/yellow opponent signals, respectively (Figure 15).

There are small systematic differences in color statistics between animal and non-animal images in our database. Animal images tend to have more mid-range luminance and fewer black and bright pixels than non-animal images. Animal

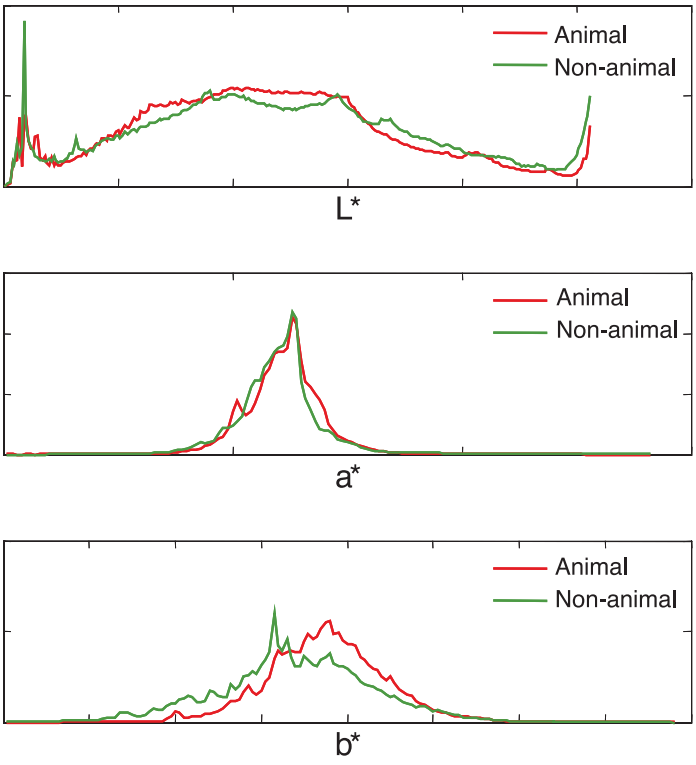

Figure 15. L*a*b* color histograms for ensemble of images used in Experiment 1.

images tend to have more pixels toward magenta and yellow, while non-animal images are more biased to blue.

To assess the utility of these small color signals (when used alone) for animal detection, we evaluated a simple algorithm that categorizes an image as animal or non-animal based on the χ^2 difference between the color histogram of the image and the average animal and non-animal color histograms derived from the other images in the database. This algorithm is nearly optimal when pixels are independent. While this is of course far from the case in natural imagery, it is reasonable here to ignore spatial dependencies between pixels since we wish to evaluate the information in color independent of spatial cues. However, non-spatial dependencies no doubt also exist, and the simple χ^2 algorithm ignores these. Thus, the performance of the χ^2 algorithm should be considered a lower bound on the detection performance possible when discrimination is based solely on color and luminance.

We tested a number of color spaces and histogram sampling rates and found best performance on our database using the L*a*b* color space and 256 bins per channel: 76% correct ($d' = 1.41$ for an unbiased participant). This performance is well over the maximum contribution for color and luminance alone estimated from our experiments ($d' = .3$; Figure 5) but well under the performance we witnessed for our participants on complete natural image stimuli ($d' > 3.5$).

Interestingly, ignoring the luminance channel and basing discrimination only on the color-opponent a* and b* channels yielded nearly identical performance (75% correct, $d' = 1.35$). Conversely, discrimination using only the luminance channel L* yielded lower performance (65%, $d' = .77$).

These results suggest that both luminance and color provide potentially useful cues on their own (i.e., without shape or spatial-layout information), particularly since the above results must be considered a lower bound. Yet the human visual system does not appear to exploit these cues, at least for rapid animal detection.

One consideration is that the images used here, obtained from a stock commercial database, may not be representative of images we would encounter when free viewing and moving naturally through real scenes. Beyond potential differences in color transformation, these photographs have been composed with specific artistic intentions, and the animal scenes have presumably been photographed after the animals were already detected by the photographer's eye. In some cases, the animals themselves may be posed.

Nevertheless, these results suggest that it may not be enough to assume that color is not used for animal detection because it is not diagnostic (Humphrey et al., 1994; Tanaka & Presnell, 1999; Wurm et al., 1993). We must ask why, if the signal exists, it is not used.

We stress that the color signals observed in Figure 15 derive from entire images: It is not clear what proportion of these signals derives from the animal itself and what proportion from the background. Sensitivity to color signals from the animal itself may be low if the evolution and development of our visual system has been driven more by detection of potential predators and prey whose colors are typically well matched to background. Sensitivity to background color signals may be low if our visual system is driven by the goal of localizing the animal within the image, not just detecting presence.

On the other hand, shape and texture cues from the animal provide a direct means for detection regardless of color camouflage and can also provide rapid information on 3D spatial layout, which may provide an indirect basis for detection (Oliva & Torralba, 2007).

Models for rapid object detection

Given the speed with which animals are detected in natural scenes, it is natural to consider the possibility that rapid object detection is a saliency phenomenon: Perhaps detection is rapid because the animals serve as powerful exogenous cues that cause them to perceptually "pop-out" from their background.

Early computational models of image salience (Itti, Koch, & Niebur, 1998) are based on simple local measures of luminance, color, and orientation contrast, all shown to be statistically significant predictors of human attention (Parkhurst, Law, & Niebur, 2002). For some of our animal images, the animal do stand out based upon luminance, color, and texture features. But for others, the animals are fairly well camouflaged (Figure 2).

Interestingly, more recent studies (Mundhenk & Itti, 2005; Peters, Iyer, Itti, & Koch, 2005) have shown that incorporation of more complex short- and long-range contour interactions improves the predictive power of salience models. Similar mechanisms might be involved in computing shape cues on which rapid animal detection is based. Since long-range interactions require greater computation, they may be involved only for longer SOAs.

Again, we do not know from our results, or from previous experiments on animal detection, to what degree detection is based upon features of the animal per se and to what degree upon features of the scene context. If the latter, this would suggest that rapid animal detection has less to do with object pop-out than rapid scene analysis.

Visually salient features are assumed to draw attention. Could attention be involved in the rapid detection of animals in natural scenes? The very limited processing time available to account for the most rapid responses seems insufficient to allow both redirection and full application of attentional resources. There are a number of studies that support this view. Rousselet, Fabre-Thorpe, and Thorpe (2002) found that detection was as fast for two images presented simultaneously as for a single image. Li, VanRullen, Koch, and Peron (2002) found no decrement in performance when participants performed a simultaneous secondary attention task. Evans and Treisman (2005) found a dissociation between animal detection and identification in an RSVP task, suggesting that

detection might be relying upon more primitive mechanisms. Further, they found that detection was largely immune to the attentional blink phenomenon, providing further evidence that rapid detection may be based largely upon pre-attentive mechanisms. Based on these and other results, they have suggested that rapid animal detection may be based upon the feedforward detection of a disjunctive set of features diagnostic of the target category, without actual binding of these features into a coherent representation sufficient for identification.

In the context of our findings, this proposal would suggest that rapid performance is based upon the feedforward detection of diagnostic texture patches and shape fragments, without requiring complete linking of bounding contours into closed forms. This proposal is consistent with traditional accounts of pop-out phenomena (Treisman & Gelade, 1980) and with more recent models and computer vision algorithms for feedforward object detection in natural scenes (Serre, Wolf, Bileschi, Riesenhuber, & Poggio, 2007; Shotton, Blake, & Cipolla, 2008; Ullman, Vidal-Naquet, & Sali, 2002).

For example, the HMAX model proposed by Serre et al. (2007) bases discrimination upon weighted combinations of roughly linear fragments, each of which are partially position and scale invariant and corresponding roughly to shape and texture cues as defined in our experiments. Serre et al. evaluate the model on a rapid animal detection task and find that the model performance correlates well with human performance when evaluated on various kinds of animal scenes.

These theories make a specific prediction: Scrambling natural scenes while preserving texture and shape features should have no effect on rapid object detection. Recent experiments by Hayworth, Yue, and Biederman (2007) show that human judgments of scrambled line drawings are not consistent with the HMAX theory of Serre et al. (2007). However, to our knowledge, a test of the effects of scrambling on rapid animal detection in briefly presented natural scenes has not been reported.

While the most rapid detections may suggest purely feedforward feature-based discrimination, the continued improvement in performance with longer stimulus durations seen in our experiments and the broad distribution in the onset of ERP differentials correlated with longer reaction times (Johnson & Olshausen, 2003) suggest a role for recurrent and/or attentional processes in object detection (Cavanagh, 1991; Di Lollo, Enns, & Rensink, 2000; Lee & Mumford, 2003; Rao & Ballard, 1999; Treisman & Gelade, 1980).

Conclusion

In this study, we have systematically manipulated natural imagery to estimate the respective role and dynamics of proximal cues to the presence of animals in natural scenes. Despite carrying useful information for the task, color and luminance cues were found to have at most a minor impact on detection. Rather, detection appears to be largely determined by shape and texture cues. Interestingly, shape cues appear to be extracted more efficiently than texture cues, requiring only 12–17 msec of backward-masked stimulus exposure to yield information relevant to the task. Detection is as fast for edge map stimuli produced using simple automatic filtering techniques as it is for human-generated outlines, suggesting that these shape cues could be extracted efficiently from natural images by simple neural mechanisms in early visual cortex.

Acknowledgments

We thank Bob Hou for programming support and the two anonymous reviewers for their helpful comments. This work was supported by research grants from NSERC, GEOIDE, and OCE.

Commercial relationships: none.
Corresponding author: James H. Elder.
Email: jelder@yorku.ca.
Address: Centre for Vision Research, 0009 CSEB, York University, 4700 Keele Street, Toronto, Ontario, Canada M3J 1P3.

References

Averbach, E., & Coriell, A. S. (1961). Short-term memory in vision. *Bell System Technical Journal, 40,* 309–328.

Biederman, I. (1972). Perceiving real-world scenes. *Science, 177,* 77–80. [PubMed]

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147. [PubMed]

Biederman, I., Glass, A. L., & Stacey, E. W., Jr. (1973). Searching for objects in real-world scenes. *Journal of Experimental Psychology, 97,* 22–27. [PubMed]

Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology, 20,* 38–64. [PubMed]

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 443–446. [PubMed]

Brodie, E. E., Wallace, A. M., & Sharrat, B. (1991). Effects of surface characteristics and style of production on naming and verification of pictorial stimuli. *American Journal of Psychology, 104,* 517–545. [PubMed]

Bruner, J. (1957). *Contemporary approaches to cognition*. Cambridge: Harvard University Press.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 8,* 679–698.

Cavanagh, P. (1991). What's up in top-down processing? In A. Gorea (Ed.), *Representations of vision: Trends and tacit assumptions in vision research* (pp. 295–304). Cambridge, UK: Cambridge University Press.

Davidoff, J. B., & Ostergaard, A. L. (1988). The role of color in categorical judgements. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 40,* 533–544. [PubMed]

Delorme, A., Richard, G., & Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: A study in monkeys and human. *Vision Research, 40,* 2187–2200. [PubMed]

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General, 129,* 481–507. [PubMed]

Elder, J. H., Krupnik, A., & Johnston, L. A. (2003). Contour grouping with prior models. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 25,* 661–674.

Elder, J. H., & Zucker, S. W. (1998). Local scale control for edge detection and blur estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20,* 699–716.

Epstein, R., Graham, K. S., & Downing, P. E. (2003). Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron, 37,* 865–876. [PubMed]

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392,* 598–601. [PubMed]

Evans, K. K., & Treisman, A. (2005). Perception of objects in natural scenes: Is it really attention free? *Journal of Experimental Psychology: Human Perception and Performance, 31,* 1476–1492. [PubMed]

Fei-Fei, L., VanRullen, R., Koch, C., & Perona, P. (2005). Why does natural scene categorization require little attenion? Exploring attentional requirements for natural and synthetic stimuli. *Visual Cognition, 12*.

Goffaux, V., Jacques, C., Mouraux, A., Oliva, A., Schyns, P. G., & Rossion, B. (2005). Diagnostic colours contribute to the early stages of scene categorization: Behavioral and neuropsychological evidence. *Visual Cognition, 12,* 878–892.

Hayworth, K., Yue, X., & Biederman, I. (2007). Some tests of the standard model [Abstract]. *Journal of Vision, 7*(9):924, 924a, http://journalofvision.org/7/9/924/, doi:10.1167/7.9.924.

Humphrey, G. K., Goodale, M. A., Jakobson, L. S., & Servos, P. (1994). The role of surface information in object recognition: Studies of a visual form agnosic and normal subjects. *Perception, 23,* 1457–1481. [PubMed]

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20,* 1254–1259.

Johnson, J. S., & Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. *Journal of Vision, 3*(7):4, 499–512, http://journalofvision.org/3/7/4/, doi:10.1167/3.7.4. [PubMed] [Article]

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research, 46,* 1762–1776. [PubMed]

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 20,* 1434–1448. [PubMed]

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America, 99,* 9596–9601. [PubMed] [Article]

Martin, D., Fowlkes, C., & Malik, J. (2004). Learning to detect natural image boundaries using local brightness, color and texture cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 26,* 530–549. [PubMed]

Mundhenk, T. N., & Itti, L. (2005). Computational modeling and exploration of contour integration for visual saliency. *Biological Cybernetics, 93,* 188–212. [PubMed]

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology, 41,* 176–210. [PubMed]

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends and Cognitive Science, 11,* 520–527. [PubMed]

Ostergaard, A. L., & Davidoff, J. B. (1985). Some effects of color on naming and recognition of objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 579–587. [PubMed]

Parent, P., & Zucker, S. W. (1989). Trace inference, curvature consistency, and curve detection. *IEEE*

Transactions on Pattern Analysis and Machine Intelligence, 11, 823–839.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42,* 107–123. [PubMed]

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442. [PubMed]

Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research, 45,* 2397–2416. [PubMed]

Price, C. J., & Humphrey, G. W. (1989). The effects of surface detail on object recognition of objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 579–587.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2,* 79–87. [PubMed]

Renninger, L. W., & Malik, J. (2004). When is scene identification just texture recognition? *Vision Research, 44,* 2301–2311. [PubMed]

Roberts, W. A., & Mazmanian, D. S. (1988). Concept learning at different levels of abstraction by pigeons, monkeys, and people. *Journal of Experimental Psychology: Animal Behavior Processes, 14,* 247–260.

Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience, 5,* 629–630. [PubMed]

Sanocki, T., Bowyer, K. W., Heath, M. D., & Sarkar, S. (1998). Are edges sufficient for object recognition? *Journal of Experimental Psychology: Human Perception and Performance, 24,* 340–349.

Schiller, P. H., Logothetis, N. K., & Charles, E. R. (1991). Parallel pathways in the visual system: Their role in perception at isoluminance. *Neuropsychologia, 29,* 433–441. [PubMed]

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 6424–6429. [PubMed] [Article]

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., & Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29,* 411–426. [PubMed]

Sha'ashua, A., & Ullman, S. (1988). *Structural saliency: The detection of globally salient structures using a locally connected network*. Paper presented at the Proceedings of the 2nd International Conference on Computer Vision.

Shotton, J., Blake, A., & Cipolla, R. (2008). Multi-scale categorical object recognition using contour fragments. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 30,* 1270–1281. [PubMed]

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74,* 1–29.

Steeves, J. K., Humphrey, G. K., Culham, J. C., Menon, R. S., Milner, A. D., & Goodale, M. A. (2004). Behavioral and neuroimaging evidence for a contribution of color and texture information to scene classification in a patient with visual form agnosia. *Journal of Cognitive Neuroscience, 16,* 955–965. [PubMed]

Tanaka, J. W., & Presnell, L. M. (1999). Color diagnosticity in object recognition. *Perception & Psychophysics, 61,* 1140–1153. [PubMed]

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381,* 520–522. [PubMed]

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136. [PubMed]

Ullman, S. (1989). Aligning pictorial description: An approach to object recognition. *Cognition, 32,* 193–254. [PubMed]

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience, 5,* 682–687. [PubMed]

VanRullen, R., & Thorpe, S. J. (2001a). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. *Perception, 30,* 655–668. [PubMed]

VanRullen, R., & Thorpe, S. J. (2001b). The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience, 13,* 454–461. [PubMed]

Vogel, R. (1999a). Categorization of complex visual images by rhesus monkeys. Part 1: Behavioural study. 1223–1238. [PubMed]

Vogel, R. (1999b). Categorization of complex visual images by rhesus monkeys. Part 2: Single-cell study. *European Journal of Neuroscience, 11,* 1239–1255. [PubMed]

Williams, L. R., & Thornber, K. K. (1999). A Comparison of measures for detecting natural shapes in cluttered backgrounds. *International Journal of Computer Vision, 34,* 81–96.

Wurm, L. H., Legge, G. E., Isenberg, L. M., & Luebker, A. (1993). Color improves object recognition in normal and low vision. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 899–911. [PubMed]