Running Head: AREA ACTIVATION

Advancing Area Activation towards a General Model of Eye Movements in Visual Search

Marc Pomplun

Department of Computer Science

University of Massachusetts at Boston

100 Morrissey Boulevard

Boston, MA 02125-3393

USA

Phone: 617-287-6485

Fax: 617-287-6433

e-mail: marc@cs.umb.edu

Abstract

Many of our everyday tasks require us to perform visual search. Therefore, an adequate model of visual search is an indispensable part of any plausible approach to modeling integrated cognitive systems that process visual input. Due to its quantitative nature, absence of freely adjustable parameters, and support from empirical research results, the Area Activation Model is presented as a promising starting point for developing such a model. Its basic assumption is that eye movements in visual search tasks tend to target display areas that provide a maximum amount of task-relevant information for processing. To tackle the shortcomings of the current model, an empirical study is briefly reported which provides a variety of quantitative data on saccadic selectivity in visual search. It is discussed how these and related data will be used to develop the Area Activation Model towards a general model of eye movements in visual search.

Advancing Area Activation towards a General Model of Eye Movements in Visual Search

1. Introduction

During every single day of our lives we perform thousands of elementary tasks, which often are so small and require so little effort that we hardly even recognize them as tasks. Think of a car driver looking for a parking spot, a gardener determining the next branch to be pruned, a painter searching for a certain color on her palette, or an Internet surfer scrutinizing a web page for a specific link. These and other routine tasks require us to perform visual search. Given this ubiquity of visual search, it is not surprising that we excel at it. Without great effort, human observers clearly outperform every current artificial vision system in tasks such as finding a particular face in a crowd or determining the location of a designated item on a desk. Understanding the mechanisms underlying visual search behavior will thus not only shed light on crucial elementary functions of the visual system, but it may also enable us to devise more efficient and more sophisticated computer vision algorithms. Moreover, an adequate model of visual search is an indispensable part of any plausible approach to modeling integrated cognitive systems that consider visual input. As a consequence, for several decades visual search has been one of the most thoroughly studied paradigms in vision research.

In a visual search task, subjects usually have to decide as quickly and as accurately as possible whether a visual display, composed of multiple search items, contains a pre-specified target item. Many of these studies analyzed the dependence of response times and error rates on the number of search items in the display. Although rather sparse, such data led to the development of numerous theories of visual search. These theories differ most significantly in the function they ascribe to visual attention and its control in the search process. For an introduction to the questions and approaches in the field of visual search see Jeremy Wolfe's chapter in the present volume. The same author also wrote a comprehensive review on visual search (Wolfe, 1998).

Furthermore, it was Jeremy Wolfe and his colleagues who proposed one of the most influential theories of visual search, named the Guided Search Theory (e.g., Cave & Wolfe, 1990; Wolfe 1994, 1996; Wolfe, Cave, & Franzel, 1989; see also Wolfe's chapter in this volume). The basic idea underlying this theory is that visual search consists of two consecutive stages: an initial stage of preattentive processing that guides a subsequent stage of serial search. After the onset of a search display, a parallel analysis is carried out across all search items, and preattentive information is derived to generate an "activation map" that indicates likely target locations. The overall activation at each stimulus location consists of a top-down and a bottom-up component. A search item's top-down (goal-driven) activation increases with greater similarity of that item to the target, whereas its bottom-up (data-driven) activation increases with decreasing similarity to other items in its neighborhood. This activation map is used to guide shifts of attention during the subsequent serial search process. First, the subject's focus of attention is drawn to the stimulus location with the highest activity. If the target actually is at this location, the subject manually reports target detection, and the search trial terminates. Otherwise, the subject's attention moves on to the second-highest peak in the activation map, and so on, until the subject either detects the target or decides that the display does not contain a target.

The Guided Search Theory has been shown to be consistent with a wide variety of psychophysical visual search data (e.g., Brogan, Gale, & Carr, 1993). Besides the standard measures of response time and error rate, these data also encompass more fine-grained measures, most importantly eye-movement patterns. In static scenes such as standard search displays, eye movements are performed as alternating sequences of saccades (quick "jumps", about 30-70 ms) and fixations (almost motionless phases, about 150-800 ms). Interestingly, information from the display is extracted almost entirely during fixations (for a review of eye-movement research see Rayner, 1998). Therefore, the positions of fixations - or saccadic endpoints – can tell us which display items subjects looked at during a visual search trial before they determined the presence or absence of the target. Analyzing the features of the inspected items and relating them to the features of the target item can provide valuable insight into the search process. Based on this idea, several visual search studies have examined saccadic selectivity, which is defined as the proportion of saccades directed to each type of non-target item (distractor), by assigning each saccadic endpoint to the nearest item in the search display. The Guided Search Theory received support from several of these studies which revealed that those distractors sharing a certain feature such as color or shape with the target item received a disproportionately large number of saccadic endpoints (e.g., Findlay, 1997; Hooge & Erkelens, 1999; Motter & Belky, 1998; Pomplun, Reingold, & Shen, 2001b; Scialfa & Joffe, 1998; Shen, Reingold, & Pomplun, 2000; Williams & Reingold, 2001; but see Zelinsky, 1996).

At this point, however, a closer look at the relationship between eye movements and visual attention is advisable. In the previous paragraph we implicitly assumed that the items subjects look at are also the ones that receive their attention. From our everyday experience we know that this does not always have to be correct: First, we can direct our gaze to an object in our visual field without paying attention to the object or inspecting it – we could simply think of something completely unrelated to the visual scene, maybe an old friend we have not seen in years. Second, even if we inspect the visual scene, we are able to process both the item that we are currently fixating and its neighboring items. For example, when fixating on any of the bar items in Figure 1a, we can sequentially shift our attention to each of its neighboring items, without moving our eyes, and thereby determine their brightness and orientation. These covert shifts of attention work efficiently for items near the fixation, but become less feasible with increasing retinal eccentricity; for instance, while fixating on the center of Figure 1a, we cannot examine any specific item in Figure 1b (see Rayner, 1998). It has been shown by several studies that subjects typically process multiple items within a single fixation during visual search tasks (e.g., Bertera & Rayner, 2000; Pomplun, Reingold & Shen, 2001a).

The first phenomenon – inattention - can be accounted for reasonably well by asking subjects to perform the visual search task as quickly and as accurately as possible and only analyzing those trials in which subjects gave a correct manual response within three standard deviations from the mean response time. The second phenomenon - covert shifts of attention - however, is inherent to eye-movement research: It is impossible to infer from a subject's gaze trajectory the exact sequence of items that were processed. We can only estimate this sequence, as saccades roughly follow the focus of attention to provide high visual acuity for efficient task performance. So it is important to notice that the nearest-item definition of saccadic selectivity, while it can identify features that guide search, does not measure attentional selectivity, that is, the proportion of attention directed to each distractor type.

Undoubtedly, modeling visual attention through Guided Search has already advanced the field of visual search quite substantially – so what would be the additional benefit of quantitatively predicting the position and selectivity of saccadic endpoints? First of all, quantitative modeling is important for the evaluation and comparison of different models. Since eye movements - unlike shifts of attention - can be directly measured, the empirical testing of models predicting eye movement data is especially fruitful. Moreover, if we want to integrate our model into an embodied computational cognitive architecture, the model needs to accept quantitative input and produce quantitative output in order to interact with the other components in the architecture. The most important argument for the quantitative modeling of eye movements, however, is the fact that for executing elementary tasks we use our gaze as a pointer into visual space. We thereby create an external coordinate system for that particular task that greatly reduces working memory demands (see Ballard, 1991; Ballard, Hayhoe, Pook, & Rao, 1997). For instance, the task of grasping an object is typically performed in two successive steps: First, we fixate on the object, and second, we move our hand towards the origin of the gaze-centered coordinate system (Milner & Goodale, 1995). Since visual search is a component of so many elementary tasks, having a quantitative model of eye-movement control in visual search is crucial for simulating and fully understanding the interaction of small-scale processes that enable us to efficiently perform a large variety of natural tasks. Such a model could serve as a valuable visual search module for the more ambitious endeavor of accurately modeling integrated cognitive systems.

Rao, Zelinsky, Hayhoe, and Ballard (2001) propose a computational eye-movement model for visual search tasks that uses iconic scene representations derived from oriented spatiochromatic filters at multiple scales. In this model, visual search for a target item proceeds in a coarse-to-fine fashion. The first saccadic endpoint is determined based on the target's largest scale filter responses, and subsequent endpoints are based on filters of decreasing size. This coarse-to-fine processing, which is supported by psychophysical data (e.g., Schyns & Oliva, 1994), makes this model a biologically plausible and intuitive approach. It is important to notice, however, that the psychophysical data also show the duration of the coarse-to-fine transition not to exceed a few hundred milliseconds after stimulus onset, which makes it especially relevant to short search processes. Accordingly, the model was tested on a visual search task that was easier, and therefore shorter, than typical tasks in the literature. It will be interesting to see the further development of this model towards a greater variety of search tasks.

A very simple, quantitative approach to modeling the spatial distribution and feature selectivity of eye movements in longer visual search tasks is the Area Activation Model (Pomplun, Reingold, Shen, & Williams, 2000; Pomplun, Shen, & Reingold, 2003). Its original version only applies to artificial search displays with discrete search items, and it requires substantial a-priori information about the guiding features and the task difficulty in order to work accurately. Nevertheless, the model made novel predictions that were successfully tested in an empirical study. Due to its straightforward nature and ability of precise prediction, the Area Activation Model - once its restrictions have been tackled – can be considered a promising candidate for a general visual search component for integrated models of cognitive systems. The following Section 2 will briefly introduce the original Area Activation Model, while Section 3 will describe current efforts and preliminary results regarding the improvement of the model towards a general visual search module.

2. The Original Area Activation Model

The original Area Activation Model (Pomplun et al., 2000; 2003) is related to the Guided Search Theory as it also assumes a preattentively generated activation map that determines feature guidance in the subsequent search process. The most important difference between the two models is the functional role of the activation map. Guided Search assumes an activation map containing a single peak for each relevant search item; this map is used to guide visual attention. Area Activation proposes an activation map containing for every position in the search display the amount of relevant information that could be processed during a fixation at that position; this map determines the position of fixations. To be precise, the Area Activation Model is based on assumptions concerning three aspects of visual search performance: first, the extent of available resources for processing; second, the choice of fixation positions; and third, the scan-path structure. These assumptions will be briefly described in the following paragraphs.

Regarding the resources available for visual processing, it is assumed that during a fixation the distribution of these resources on the display can be approximated by a two-dimensional Gaussian function centered at the fixation point (cf. Pomplun, Ritter, & Velichkovsky, 1996). The region in the display covered by this distribution is also called the fixation field. We define it as the area from which task-relevant information is extracted during a fixation. The size of this area – technically speaking, the standard deviation of the Gaussian function - depends upon numerous stimulus dimensions such as task difficulty, item density, and item heterogeneity. For example, in displays with widely dispersed items, where the distractor items are clearly different from the target, the fixation field is expected to be larger than in displays with densely packed items and distractors that are very similar to the target (see Bertera & Rayner, 2000). The model further assumes that a smaller fixation field requires more fixations to process a search display. This is quite intuitive, because if a smaller area can be processed with each fixation, more fixations are needed to process the entire display. Such a functional relationship makes it possible to estimate the fixation field size based on the empirically observed number of fixations in a particular experimental condition. Consequently, the model must be given an estimate of the number of fixations that subjects will generate, ideally obtained through a pilot study. An iterative gradient-descent algorithm is used on a trial-by-trial basis to determine the fixation field size in such a way that the number of simulated fixations matches the number of empirical ones.

Concerning fixation positioning, the model assumes that fixations are distributed in such a way that each fixation processes a local maximum of task-relevant information. To determine these positions, the model first computes the informativeness of positions across the display, that is, the amount of task-relevant information that can be processed during a fixation at that position. The model calculates informativeness of a fixation position as the sum of visual processing resources – according to the Gaussian function centered at that position – that is applied to each display item having one or more of the guiding features (these items will be referred to as guiding items). In other words, the value of the Gaussian function for the position of each of guiding items is determined, and the sum of these values equals the informativeness. This means that positions in the display within dense groups of guiding items are most informative, because if observers fixate on those positions, they can process many guiding items and are more likely to find the target than during a fixation on a less informative position. Computing the informativeness for every pixel in a search display results in a smooth activation function that represents the saccade-guiding activation map of the Area Activation Model (see the examples in Figures 1c and 1d). If the fixation field is sufficiently large, local groups of guiding search items will induce a single activation peak in their center. Consequently, the model will predict a fixation in the center of this group to be more likely than a fixation on any individual guiding item. This center-of-gravity effect has been empirically observed in numerous studies (e.g., Findlay, 1997; Viviani & Swensson, 1982). Once the activation map has been generated, the choice of fixation positions by the model is a statistical process: More highly activated - that is, more informative – positions are more likely to be fixation targets than less activated positions. Notice that in order to perform all of these computations the model must be "told" which feature or features of search items guide the search process, which can be determined through a pilot study.

Finally, to compute a visual scan path, the model not only needs to determine the fixation positions, but also the order in which they are visited. As indicated by empirical studies (e.g., Zelinsky, 1996), determination of such an order is geared towards minimizing the length of the scan path. The following simple rule in the Area Activation Model reflects this principle: The next fixation target is always the novel activation peak closest to the current gaze position. This method of local minimization of scan path length ("Greedy Heuristic") was shown to adequately model empirical scanning sequences (e.g., Pomplun, 1998).

A complete mathematical description of the Area Activation Model is provided in Pomplun et al. (2003), and its application to another visual search task involving simultaneous guidance by multiple features is described in Pomplun et al. (2000). In the present context, however, the operation of the Area Activation Model will only be demonstrated by outlining the empirical testing of one of its major predictions. This prediction is rather counterintuitive and clearly distinct from those made by other models such as Guided Search – it states that saccadic selectivity depends on the spatial arrangement of search items in the display. In other words, according to the Area Activation Model, two displays that show identical, but differently arranged sets of items can induce substantially different patterns of saccadic selectivity as measured by the nearest-item method. This is because the peaks in the activation map do not necessarily coincide with the positions of the guiding distractors, but may have a distractor of another type as their nearest neighbor. As a consequence, a spatial arrangement of search items that leads to a closer match between the positions of activation peaks and distractors of a particular type will produce higher saccadic selectivity towards this distractor type. In contrast, models assuming a single activation peak for each guiding distractor predict a pattern of saccadic selectivity that is independent of the arrangement of search items.

To test this prediction, two groups of visual search displays were created. Although all of these displays contained the same set of search items, they were - according to the Area Activation Model - hypothesized to induce relatively strong saccadic selectivity ("high guidance displays") or relatively weak saccadic selectivity ("low guidance displays") towards the guiding distractor type. Subjects showing a significant selectivity difference in the predicted direction between these two groups of displays would yield strong support for the model.

--- insert Figure 1 about here ---

To demonstrate the model's performance, Figure 1 shows a simplified variant of the stimuli used in the original study. The search items are bars of different brightness (bright vs. dark) and orientation (vertical vs. horizontal), with a bright vertical bar serving as the target. In a preliminary study, it was shown that only brightness but not orientation guided the search process. This means that the subjects' attention was attracted by the bright horizontal bars but not by the dark vertical or dark horizontal bars. In the high-guidance display shown in Figure 1a, the search items are arranged in such a way that all the peaks in the activation function (see Figure 1c) computed by the model coincide with guiding items, that is, bright horizontal bars. The low-guidance display in Figure 1b, however, includes several activation peaks whose nearest neighbors are non-guiding items, that is, dark vertical bars (see Figure 1d). Figures 1e and 1f show predicted scan paths for the high- and low-guidance display, respectively, with circles indicating fixation positions and numbers marking their temporal order. Notice that the model assumes a first fixation at the center of the display and a final fixation on the detected target, neither of which is included in the computation of saccadic selectivity.

In the actual study, each of eight subjects performed the visual search task on a set of 480 displays, which was composed of 240 high-guidance and 240 low-guidance displays. The analysis of the subjects' eye-movement data revealed that their saccadic selectivity towards the guiding items in the high-guidance displays was about 33% greater than in the low-guidance displays. Moreover, these values closely matched the ones predicted by the Area Activation Model (see Pomplun et al., 2003).

These and other empirical tests (Pomplun et al., 2000) provided supporting evidence for the Area Activation Model. Its other strong points are its straightforward nature, its consistency with environmental principles, and the absence of any freely adjustable model parameters, at least in the common case of single-feature guidance. The more such parameters a model includes, the more difficult it is to assess its performance, because these parameters can be adjusted to match simulated with empirical data, even if the underlying model is inadequate.

However, it must be stated that the shortcomings of the model are just as obvious as its advantages: First, it needs a-priori information about empirical data – the number of fixations per trial and the features guiding search – before it can generate any eye-movement predictions. Second, the model assumes that only target features guide visual attention, although it is well known that conspicuous features in the display attract attention through bottom-up activation, even if these features are not shared with the target (e.g., Thompson, Bichot, & Sato, 2005). Finally, like most visual search models, the original Area Activation model can only be applied to artificial search images with discrete search items, each of which has a well-defined set of features. These shortcomings need to be dealt with in order to further develop the Area Activation Model.

3. Towards a General Model of Eye Movements in Visual Search

The first shortcoming of the original Area Activation Model, namely its need for a-priori information, is due to its inability to estimate the task difficulty or the pattern of saccadic selectivity based on the features of the target and distractor items. How could such estimates be derived? Based on visual search literature (e.g., Wolfe, 1998), we can safely assume two things: (1) Saccadic selectivity for a certain distractor type increases with greater similarity of that distractor type with the search target, and (2) task difficulty increases both with greater similarity between target and distractor items and with decreasing similarity between different distractor types in the display. However, the crucial question is: What does the word "similarity" mean in this context? Consequently, the first aim must be to derive an operational definition of similarity and quantify its influence on saccadic selectivity.

With regard to our second aim, the introduction of bottom-up effects to the model, we should quantify how different features attract attention even without being target features. Moreover, there is another effect that may involve both bottom-up and top-down influences and should be considered, namely the distractor-ratio effect. In the present context, this effect is best explained by the eye-movement study by Shen et al. (2000), who had subjects detect a target item among two types of distractors, each of which shared a different feature with the target. While the total number of search items was held constant, the ratio between the two distractor types was varied across trials (see Figure 2 for sample stimuli and gaze trajectories taken from a related study). Saccadic selectivity towards a particular feature (brightness or orientation) was found to increase with fewer display items sharing this feature with the target, indicating that participants tended to search along the stimulus dimension shared by fewer distractors (e.g., brightness with few same-brightness distractors and orientation with few same-orientation distractors). This indicated that subjects were able to change their pattern of visual guidance to take advantage of more informative dimensions, demonstrating the flexibility of the preattentive processing. An adequate model of visual search should account for this important effect.

--- insert Figure 2 about here ---

Finally, the model's restriction to artificial search displays can be tackled at the same time as the term "similarity" is quantified. Since a generalized definition of similarity is most useful, searches in both complex and real-world images must be studied. The features to be explored should not include specific shapes or patterns, but rather should be general features. More specifically, features should be selected that are known to be relevant to the early stages of the human visual processing hierarchy such as intensity, contrast, spatial frequency, and orientation. The Area Activation concept can easily be applied to these features and their continuous distribution instead of categorical features associated with discrete items. To compute bottom-up activation, we do not need discrete search items either, but can base this computation solely on the features in the image. For instance, on average we might expect a display region of high contrast to receive more attention than a no-contrast ("empty") region, just because of its greater information content. The proportion of features in the search display can also be used to account for a continuous-image equivalent of the distractor-ratio effect, if it exists.

To address these issues, a visual search study on complex images was conducted, whose details are described in Pomplun (under review). In this study, each of 16 subjects performed 200 visual search trials. Of the 200 search displays, 120 contained real-world images that were randomly rotated by 90°, 180°, or 270° to prevent subjects from applying context-based search and trained scanning patterns (see Figure 3a for a sample display). The other 80 displays showed complex artificial images such as fractals or abstract mosaics. All images were in grayscale format using 256 gray levels. For this exploratory study of visual guidance in complex images, it was prudent to eliminate color information in order to avoid the strong attentional capture by color features. Obviously, color is an important feature that guides visual search in everyday tasks, but in order to get a better assessment of other – possibly less guiding – features, color was not included in the study described.

--- insert Figure 3 about here ---

Each of the 200 trials started with a 4-second presentation of a small 64×64 pixel image at the center of the screen. The subjects' task was to memorize this image. Subsequently, the small image was replaced by a large 800×800 pixel search image subtending a visual angle of about 25° horizontally and vertically. The subjects knew that the previously shown small image was contained somewhere in this large search display. Their task was to find the position of the small image within the large one as quickly as possible. As soon as they were sure to have found this position, subjects were to fixate on that position and press a designated button to terminate the trial. If they did not press the button within 5 seconds after the onset of the large display, the trial was ended automatically. In either case, subjects received feedback about the actual position immediately after the end of the trial.

During the search phase, the subjects' eye movements were recorded with the head-mounted SR Research EyeLink-II eye tracking system. The obtained eye-movement data made it possible to assess both the accuracy of the subjects' visual search performance and - most importantly – their saccadic selectivity. For the saccadic selectivity analysis, the fixation positions generated by all 16 subjects were accumulated for each search display. Subsequently, the distribution of visual processing across each display was calculated as follows: Every fixation in the display was associated with a Gaussian function centered at the fixation position. The maximum value of the Gaussian function was proportional to the duration of its associated fixation, and its standard deviation was one degree of visual angle. This value was chosen to match the approximate angle subtended by the human fovea. Although the Area Activation Model assumes a variable fixation field size (see Section 2), at this point a constant size had to be used because no estimate for the actual size could yet be computed. Finally, all Gaussian functions for the same display were summed, resulting in a smooth function that indicated the amount of visual processing across positions in the display. Figure 3b illustrates this function for the sample stimulus in Figure 3a. The more processing a local region in the image received, the more strongly it is overlaid with the color purple. As can clearly be seen, those regions in the image that attract most saccadic endpoints are very similar to the target area in that they contain leaves of comparable size.

The next step was to define appropriate basic image features for the analysis of saccadic selectivity. At the current state of this research, features along four dimensions have been used: intensity (average brightness of local pixels), contrast (standard deviation of local brightness), dominant spatial frequency (most elevated frequency interval in a local area as compared to baseline data), and preferred orientation (angle of dominant orientation of local edges). For a mathematical definition of these variables see Pomplun (under review). The values along each dimension were scaled to range from 0 to 1, and divided into 20 same-size intervals. Henceforth, the word "feature" will refer to one such interval within a given dimension, for example, brightness-3 or contrast-17. The visual search targets were chosen in such a way that their features varied along the full range of all dimensions.

Due to space limitations, only the analysis of contrast will be described below. The other dimensions showed similar functional behavior. Figure 3c visualizes the contrast values computed across the sample stimulus shown in Figure 3a. The more pronounced the color green is at a point in the image, the greater is the local contrast value. Notice the square at the target position containing no contrast information. In order to avoid artifacts in the analysis of saccadic selectivity, no feature information was computed or analyzed near the target positions. The reason for this is that, whenever subjects detect a target, they look at it for a certain duration before terminating the trial. So if their fixations on the target area were included in our selectivity analysis, we would find an elevated number of saccadic endpoints aimed at the target features, indicating visual guidance towards those features, regardless of whether such guidance actually exists during the search process.

The first question we can now ask is whether there is a contrast-related bottom-up effect: Do certain contrast features attract more saccadic endpoints than others, independent of the contrast in the target area? To find out, we can analyze the average amount of processing as a function of the local contrast across all displays and subjects. The result is shown in Figure 4a: Clearly, regions of high contrast (greater than 0.6) receive more processing than areas of low contrast. Given that the contrast of the target regions was distributed in the range from 0 to 1 approximately evenly, this finding indicates a general bias in favor of high-contrast regions. This is not surprising as there usually is less – or less easily available – information in low-contrast areas. So we can state that there are preferred features, that is, feature-based bottom-up effects, in searching complex images, and they can be quantified with the method described above.

--- insert Figure 4 about here ---

The next, and of course crucial, question is whether there is also feature guidance. Increased processing of those areas that share certain features with the target would indicate this type of guidance. To investigate this, the search displays were separated into three groups, namely those with low-contrast targets (0 to 0.33), medium-contrast targets (0.34 to 0.66), and high-contrast targets (0.67 to 1). Figure 4b presents the amount of processing as a function of the local contrast relative to the values shown in Figure 4a for the three groups of displays. Clearly, for the displays with low-contrast targets, there is a bias towards processing low-contrast regions, and we find similar patterns for medium- and high-contrast displays. This observation is strong evidence for visual feature guidance in complex images, which we can now quantify for any given dimension. One possible definition of the amount of guidance exerted by a particular feature dimension is the average bias in processing across all 20 feature values of that dimension, given a target of the same feature value. For example, to compute contrast guidance, we first determine the average amount of processing for the feature contrast-1 for all trials in which the target was of contrast-1 as well. To obtain the processing bias for contrast-1, we subtract from this value the average amount of processing that contrast-1 received across all 200 trials (and thus across targets of all contrast features). Then contrast guidance is calculated as the arithmetic mean of the 20 bias values derived for contrast-1 to contrast-20.

Having obtained this operational measure of guidance, one can ask whether in complex, continuous images there is a counterpart to the distractor-ratio effect in item-based search images. This can be studied by comparing the average visual guidance for trials in which the search display contains a large proportion of the target feature in a particular dimension with those trials in which this proportion is small. In analogy to the distractor-ratio effect, the small-proportion features should receive more processing than the large-proportion ones. Figure 4c shows such an analysis for the contrast dimension. The left bar shows the guidance exerted by target features with above-average presence in the displays, whereas the right bar shows the corresponding value for target features with below-average presence. The guidance for features with above-average presence is substantially smaller than for features with below-average presence, providing evidence for a continuous counterpart to the distractor-ratio effect ("feature-ratio effect").

The incorporation of these functional relationships into the Area Activation Model is currently in progress. It is important to carefully select the most appropriate set of feature dimensions to be used in the model. By simply having a linear combination of the four dimensions intensity, contrast, spatial frequency, and orientation determine the activation function, the current version of the Area Activation Model can often – but not reliably - approximate the distribution of actual saccadic endpoints. Figure 3d illustrates the current model's prediction of this distribution in the same way as the actual distribution is shown in Figure 3b.

4. Conclusions

This chapter has presented current work on the Area Activation Model aimed at developing a quantitative model of eye-movement control in visual search tasks. Such a model would be of great scientific utility because it could be employed as a general visual search module in integrated models of cognitive systems. While its approach originated from the Guided Search Theory, the Area Activation Model and its current development presented here clearly exceed the scope of Guided Search. Most significantly, Guided Search focuses on explaining shifts of attention during search processes, and although its Version 4.0 (see Jeremy Wolfe's chapter in this volume) supports the simulation of eye movements, Guided Search largely understands eye movements as by-products of attentional shifts. In contrast, the Area Activation Model considers an observer's gaze behavior as a central part of visual task performance. The model makes quantitative predictions about the locations of saccadic endpoints in a given search display, which have been shown to closely match empirical data. Despite the current work on the Area Activation Model, it still does not nearly reach the breadth and complexity of Guided Search. However, its quantitative prediction of overt visual behavior makes Area Activation more suitable as a visual search module for integration into cognitive architectures.

Another promising candidate for development into a general visual search component is the model by Rao et al. (2001). Just like Area Activation, this model also predicts the location of saccadic endpoints based on the similarity of local image features with target features. It differs from Area Activation in its use of spatiochromatic filters at multiple scales that are applied to the visual input in a temporal coarse-to-fine sequence. This is a plausible approach for rapid search processes, and its simulation generates saccades that closely resemble empirical ones. However, during longer search processes in complex real-world scenes it seems that the coarse-to-fine mechanism is less relevant for the quantitative prediction of eye movements. At the same time, factors such as the spectrum of local display variables, the proportion of features in the display, and the difficulty of the search task become more important. All of these factors are currently not considered by the Rao et al. model but are being incorporated into the evolving Area Activation Model.

This ongoing work on the Area Activation Model addresses significant shortcomings of its original version (Pomplun et al., 2003). As a first step, an empirical visual search study on complex images was conducted and briefly reported here. Based on the data obtained, various aspects of the influence of display and target features on eye-movement patterns have been quantified. This information has been used to tackle questions unanswered by the original Area Activation Model and to develop it further towards a more generalizable visual search model. The crucial improvements so far include the elimination of required empirical a-priori information, the consideration of bottom-up activation, and the applicability of the model to search displays beyond artificial images with discrete items and features.

Future research on the Area Activation Model will focus on determining the feature dimensions to be represented. The goal of this endeavor will be the selection of a small set of dimensions that appropriately characterizes visual guidance and allows precise predictions while being simple enough to make the model transparent. Further steps in the development of the model will include the investigation of color guidance in complex displays, which was omitted in the study presented here, and the prediction of fixation field size. Ideally, the resulting model will be straightforward, consistent with natural principles, and carefully avoiding any freely adjustable model parameters to qualify it as a streamlined and general approach to eye-movement control in visual search. It will certainly be a long and challenging road towards a satisfactory model for the integration into a unified cognitive architecture, but the journey is undoubtedly worthwhile.

Acknowledgements

I would like to thank Eyal M. Reingold, Jiye Shen, and Diane E. Williams for their valuable help in devising and testing the original version of the Area Activation Model. Furthermore, I am grateful to May Wong, Zhihuan Weng, and Chengjing Hu for their contribution to the eye-movement study on complex images, and to Michelle Umali for her editorial assistance.

References

Ballard, D.H. (1991). Animate vision. Artificial Intelligence Journal, 48, 57-86.

Ballard, D.H., Hayhoe, M., Pook, P., & Rao, R. (1997). Deictic codes for the embodiment of cognition. Behavioral and Brain Sciences, 20, 723-767.

Bertera, J.H. & Rayner, K. (2000). Eye movements and the span of the effective visual stimulus in visual search. Perception & Psychophysics, 62, 576-585.

Brogan, D., Gale, A., & Carr, K. (1993). Visual search 2. London: Taylor & Francis.

Cave, K.R., & Wolfe, J.M. (1990). Modeling the role of parallel processing in visual search. Cognitive Psychology, 22, 225-271.

Findlay, J. M. (1997). Saccade target selection during visual search. Vision Research, 37, 617-631.

Hooge, I.T., & Erkelens, C.J. (1999). Peripheral vision and oculomotor control during visual search. Vision Research, 39, 1567-1575.

Milner, A.D., & Goodale, M.A. (1995). The visual brain in action. Oxford University Press, Oxford.

Motter, B.C., & Belky, E.J. (1998). The guidance of eye movements during active visual search. Vision Research, 38, 1805-1815.

Pomplun, M. (1998). Analysis and Models of Eye Movements in Comparative Visual Search. Göttingen: Cuvillier.

Pomplun, M. (under review). Saccadic selectivity in complex visual search tasks. Vision Research.

Pomplun, M., Reingold, E.M., Shen, J. & Williams, D.E. (2000). The area activation model of saccadic selectivity in visual search. In Gleitman, L.R. & Joshi, A.K. (Eds.), Proceedings of the Twenty Second Annual Conference of the Cognitive Science Society, 375-380. Mahwah, NJ: Erlbaum.

Pomplun, M., Reingold, E.M. & Shen, J. (2001a). Investigating the visual span in comparative search: The effects of task difficulty and divided attention. Cognition, 81, B57-B67.

Pomplun, M., Reingold, E.M. & Shen, J. (2001b). Peripheral and parafoveal cueing and masking effects on saccadic selectivity in a gaze-contingent window paradigm. Vision Research, 41, 2757 – 2769.

Pomplun, M., Ritter, H., & Velichkovsky, B.M. (1996). Disambiguating complex visual information: Towards communication of personal views of a scene. Perception, 25, 931-948.

Pomplun, M., Shen, J. & Reingold, E.M. (2003). Area activation: A computational model of saccadic selectivity in visual search. Cognitive Science, 27, 299 - 312.

Rao, R.P.N., Zelinsky, G.J., Hayhoe, M.M., & Ballard , D.H. (2001). Eye movements in iconic visual search. Vision Research, 42, 1447-1463.

Rayner, K. (1998). Eye Movements in reading and information processing: 20 years of research. Psychological Bulletin, 124, 372-422.

Schyns, P. G., & Oliva, A. (1994). From blobs to edges: evidence for time and spatial scale dependent scene recognition. Psychological Science, 5, 195–200.

Scialfa, C. T., & Joffe, K. (1998). Response times and eye movements in feature and conjunction search as a function of eccentricity. Perception & Psychophysics, 60, 1067-1082.

Shen, J., Reingold, E. M., & Pomplun, M. (2000). Distractor ratio influences patterns of eye movements during visual search. Perception, 29, 241-250.

Thompson, K.G., Bichot, N.P., & Sato, T.R. (2005). Frontal eye field activity before visual search errors reveals the integration of bottom-up and top-down salience. Journal of Neurophysiology, 93, 337-351

Viviani, P., & Swensson, R.G. (1982). Saccadic eye movements to peripherally discriminated visual targets. Journal of Experimental Psychology: Human Perception and Performance, 8, 113- 126.

Williams, D.E., Reingold, E.M. (2001). Preattentive guidance of eye movements during triple conjunction search tasks. Psychonomic Bulletin and Review, 8, 476-488.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. Psychonomic Bulletin & Review, 1, 202-238.

Wolfe, J. M. (1996). Extending guided search: Why guided search needs a preattentive "item map." In A.F. Kramer, M.G.H. Coles, & G.D. Logan (Eds.), Converging operations in the study of visual attention (pp. 247-270). Washington, DC: American Psychological Association.

Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), Attention (pp. 13-71). Hove, England UK.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. Journal of Experimental Psychology: Human Perception and Performance, 15, 419-433.

Zelinsky, G. J. (1996). Using eye saccades to assess the selectivity of search movements. Vision Research, 36, 2177-2187.

Figure Captions

Figure 1: Demonstration of the original Area Activation Model for displays created to induce high brightness guidance (left column) or low brightness guidance (right column). (a) and (b): sample displays with the target being a bright vertical bar; (c) and (d): activation function computed for each of the two displays; (e) and (f): predicted scan path for each display with circles marking fixation positions and numbers indicating the fixation sequence.

Figure 2: Illustration of the distractor-ratio effect on saccadic selectivity with circles marking fixation positions and numbers indicating the temporal order of fixations. The target is a bright vertical bar. (a): If the same proportion of the two distractor types is given, subjects are usually guided by brightness, i.e., they search through the bright horizontal distractors; (b): If there is a much larger proportion of bright horizontal distractors, subjects typically switch their dimension of guidance.

Figure 3: Study on saccadic selectivity in complex images. (a): search display with target area marked by a yellow square (shown to the subject as post-trial feedback); (b): distribution of saccadic endpoints during search as indicated by the amount of purple coloring; (c): local contrast in the image, with more saturated green corresponding to higher contrast; (d): distribution of saccadic endpoints as predicted by the current Area Activation Model.

Figure 4: Selected results of the saccadic selectivity study. (a): amount of processing in a display region as a function of the local contrast, indicating a bottom-up effect; (b): relative amount of processing (as compared to average values) as a function of local contrast and depending on contrast in the target area, indicating feature guidance; (c): contrast guidance (see text for definition) for above- and below-average proportion of the target-level contrast in the display, indicating a feature-ratio effect.

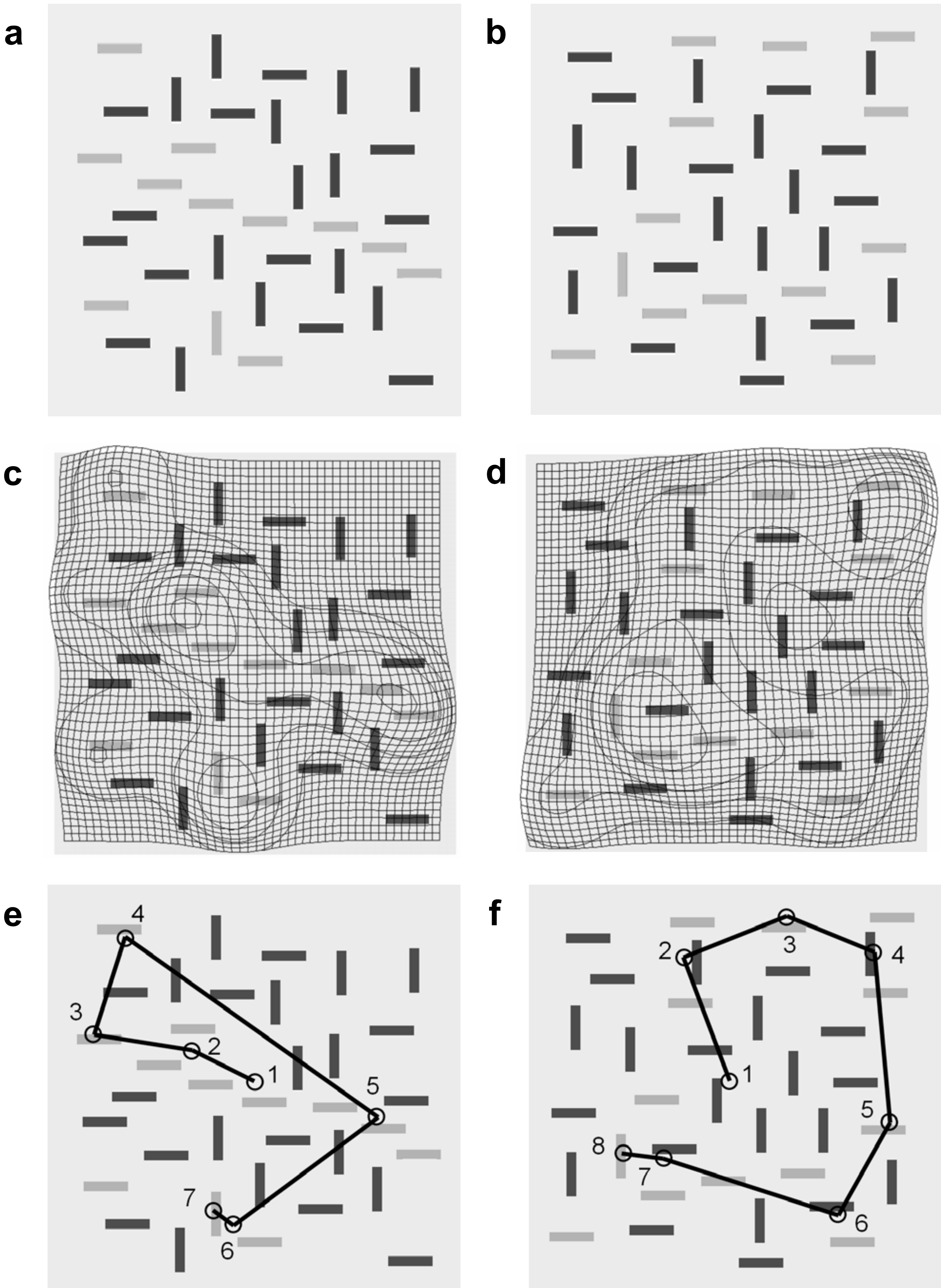

Figure 1

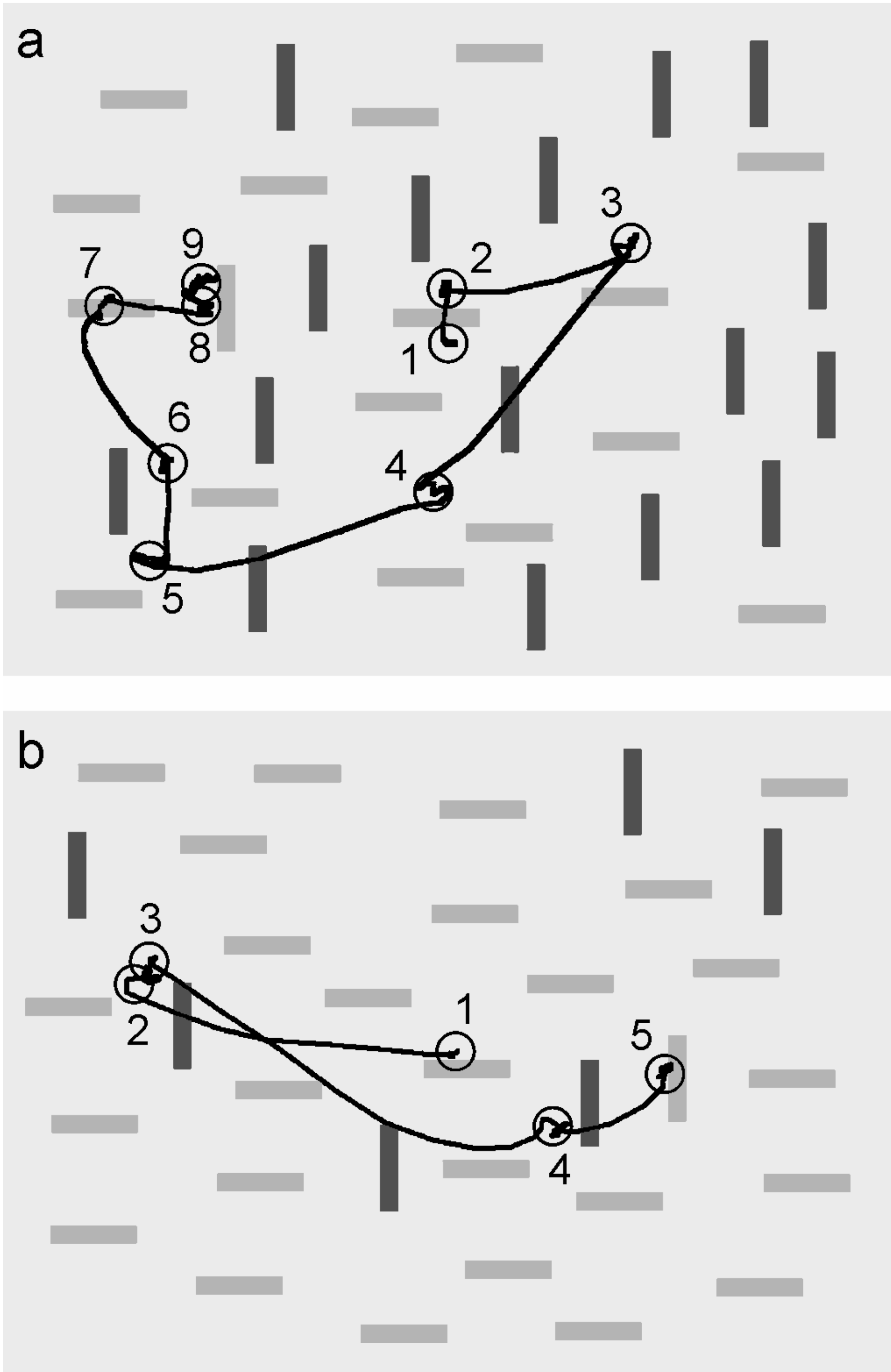

Figure 2

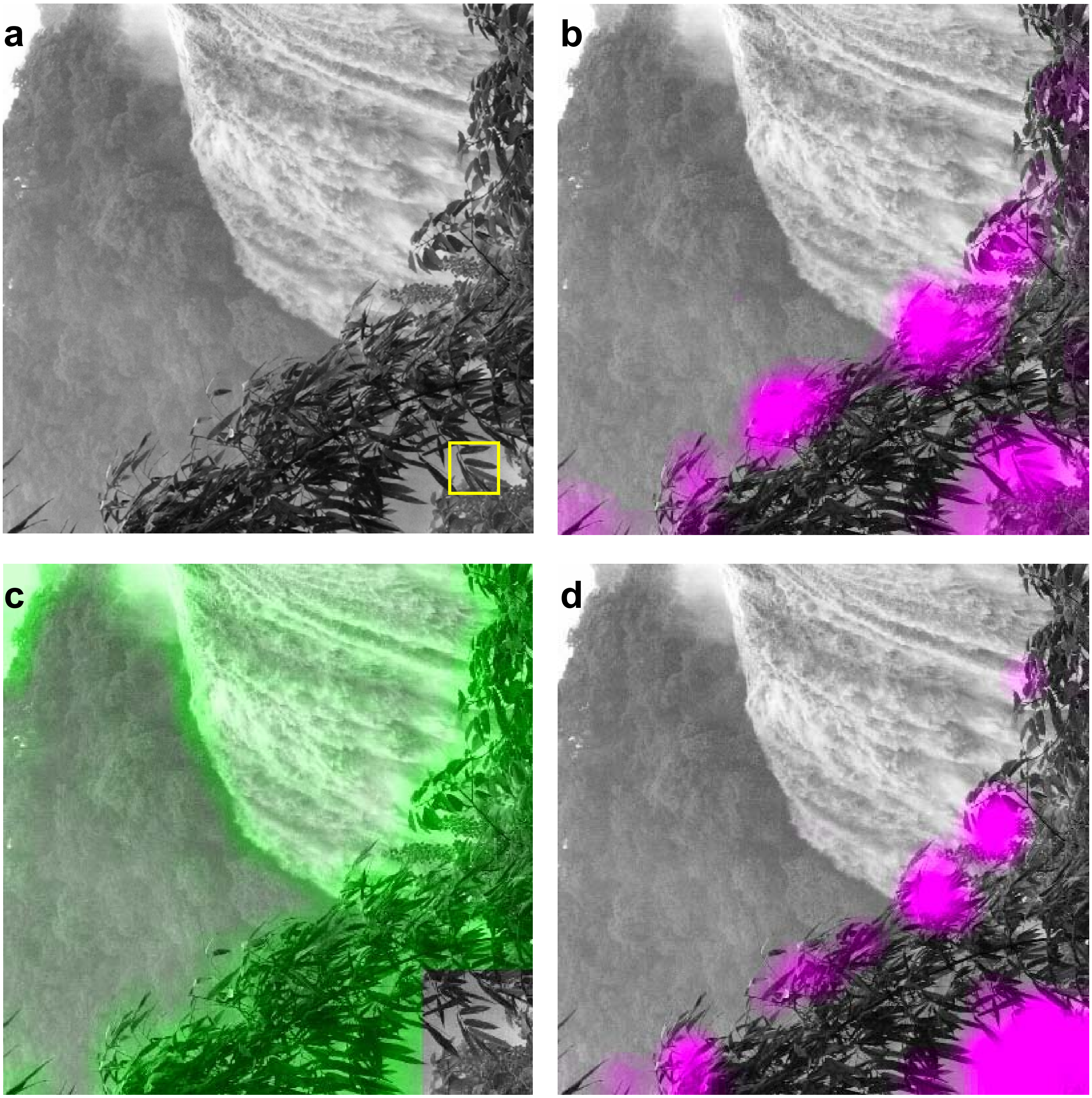

Figure 3

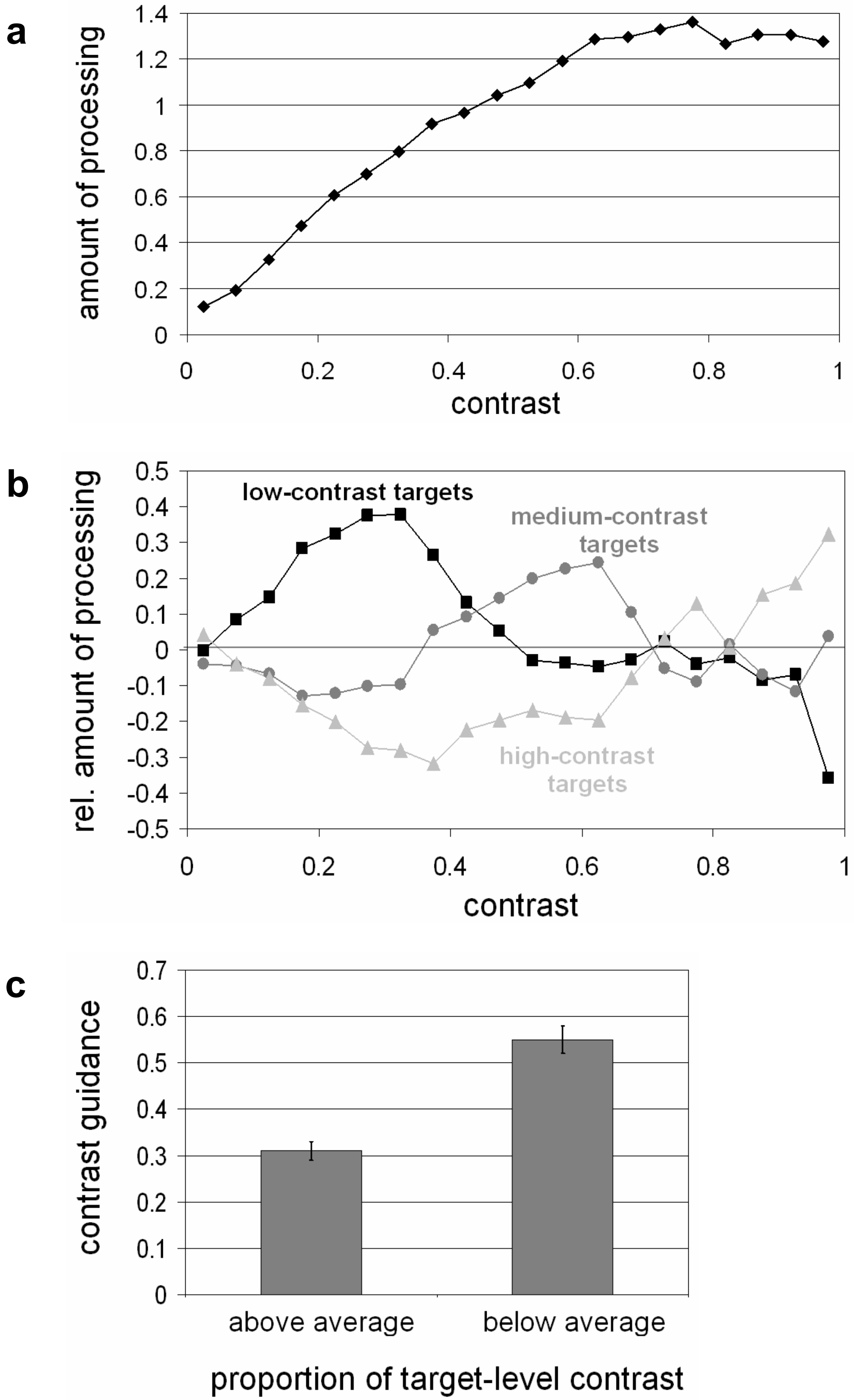
a
amount of processing
contrast
b
rel. amount of processing
low-contrast targets
medium-contrast targets
high-contrast targets
contrast
c
contrast guidance
above average
below average
proportion of target-level contrast

Figure 4