

# POSITION UNCERTAINTY, CLUTTER, AND PERFORMANCE IN NATURALISTIC SEARCH TASKS

By

YELDA SEMIZER

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Psychology

Written under the direction of

Melchi M. Michel

And approved by

---

---

---

---

New Brunswick, New Jersey

MAY, 2019

© 2019

Yelda Semizer

**ALL RIGHTS RESERVED**

## **ABSTRACT OF THE DISSERTATION**

# **Position Uncertainty, Clutter, and Performance in Naturalistic Search Tasks**

**by YELDA SEMIZER**

**Dissertation Director:**

**Melchi M. Michel**

Uncertainty regarding the position of target objects in natural scenes is a fundamental property of visual search tasks. As visual clutter in natural scenes increases, extrinsic position uncertainty (i.e., set size) also increases. While studies of visual search have repeatedly demonstrated that clutter impairs search performance in natural scenes, these studies have not attempted to disentangle the effects of search set size from those of clutter per se. Moreover, most of the clutter models used in these studies do not take the properties of the search targets into account. Thus, this dissertation has two main objectives: (1) to quantify the effect of clutter on search performance for categorical targets when the set size (i.e., extrinsic position uncertainty) is controlled and (2) to determine what visual features of categorical search targets and backgrounds are important in

measuring clutter. In Study I, we investigate the effect of natural image clutter on performance in an overt search for categorical targets when the search set size is controlled. Observers completed a search task that required detecting and localizing common objects in a set of natural images. The images were sorted into high and low clutter conditions based on the clutter metric. The search set size was varied independently, by fixing the number and positions of potential targets across set size conditions within a block of trials. Within each fixed set size condition, search times increased as a function of increasing clutter, suggesting that clutter degrades overt search performance independently of set size. In Study II, we propose new clutter metrics based on two types of target-background similarity (i.e., exemplar level and category level) to predict the effect of clutter on search performance. In a nutshell, our metrics measured the similarity between target and background features (i.e., orientation subbands) in images while also accounting for size of a search target. Our results demonstrated that both the exemplar clutter metric and the category clutter metric predicted search performance. Overall, these two studies suggest that intrinsic position uncertainty and target-background similarity should be incorporated into models of visual search and clutter in determining performance in natural search tasks.

## Acknowledgements

I would like to extend my deepest gratitude to my advisor Melchi Michel, for his continuous support, excellent guidance, and endless patience throughout the years I spent at Rutgers. I feel extremely lucky to get a chance to train under his supervision and I am grateful for all the help he has given me. I would also like to thank the rest of my committee: Eileen Kowler, Manish Singh, and Ahmed Elgammal, for their insightful suggestions regarding this work. Many thanks to all members of the Computational Vision and Psychophysics Lab and fellow graduate students of the Psychology Department for their continuous support and helpful discussions. They made my time at Rutgers more enjoyable and intellectually stimulating. I am most thankful to my family for their love, patience, and encouragement. They never stopped believing in me even when I had my doubts. I couldn't have come this far without their endless support. Finally, I would like to thank my husband Mehmet for his patience and tireless support. I am deeply thankful to him for encouraging and comforting me selflessly, especially during the last phase of my dissertation. This work was supported by NSF grant BCS-1456822.

# Table of Contents

<b>Abstract</b> . . . . .	ii
<b>Acknowledgements</b> . . . . .	iv
<b>List of Tables</b> . . . . .	ix
<b>List of Figures</b> . . . . .	x
<b>1. Introduction and Objectives</b> . . . . .	1
1.1. Introduction . . . . .	1
1.2. Objectives . . . . .	7
1.2.1. Study I: What is the effect of clutter on search performance for categorical targets when the set size is controlled? . . .	7
1.2.2. Study II: What visual features of categorical search targets and backgrounds are important in measuring clutter? . . .	8
<b>2. Visual Search and Clutter</b> . . . . .	9
2.1. Visual Search . . . . .	9
2.1.1. What Limits Search? . . . . .	9
2.1.1.1. Information Loss in the Periphery . . . . .	9
2.1.1.2. Limited Memory Capacity . . . . .	11
2.1.1.3. Limited Attentional Processing . . . . .	11
2.1.2. What Improves Search? . . . . .	12
2.1.2.1. Target Properties . . . . .	13

2.1.2.2.	Search Environment . . . . .	14
2.1.2.3.	Target Prevalence . . . . .	15
2.2.	Visual Clutter . . . . .	16
2.2.1.	How to Define Clutter? . . . . .	17
2.2.2.	How to Measure Clutter? . . . . .	18
2.2.2.1.	Clutter Measures in Psychophysics . . . . .	19
2.2.2.1.1.	The Edge Density Metric . . . . .	19
2.2.2.1.2.	The Feature Congestion Metric . . . . .	19
2.2.2.1.3.	The Subband Entropy Metric . . . . .	20
2.2.2.1.4.	The Segmentation Based Clutter Metric . . . . .	21
2.2.2.1.5.	The Proto-Object Model . . . . .	21
2.2.2.2.	Clutter Measures in Image Optics . . . . .	22
2.2.2.2.1.	The Statistical Variance Metric . . . . .	22
2.2.2.2.2.	The Probability of Edge Detection Metric . . . . .	23
2.2.2.2.3.	The Target Structure Similarity Metric . . . . .	23
2.2.2.2.4.	The Contrast Sensitivity Function Based Metric . . . . .	24
2.2.2.2.5.	The Masked Target Transform Volume Metric . . . . .	25
<b>3.</b>	<b>Study I . . . . .</b>	<b>27</b>
3.1.	Introduction . . . . .	27
3.2.	Methods . . . . .	31
3.2.1.	Observers . . . . .	31
3.2.2.	Apparatus . . . . .	32
3.2.3.	Stimuli . . . . .	32
3.2.3.1.	Creating the Image Data Set . . . . .	33

3.2.3.1.1.	Initial Filtering Stage . . . . .	33
3.2.3.1.2.	Transformation Stage . . . . .	33
3.2.3.1.3.	Labelling Stage . . . . .	34
3.2.3.1.4.	Selection Stage . . . . .	34
3.2.3.2.	Preparing the Search Stimuli . . . . .	36
3.2.3.3.	Measuring Clutter . . . . .	37
3.2.3.3.1.	Segmentation Algorithm . . . . .	37
3.2.3.3.2.	Clutter Measure . . . . .	38
3.2.4.	Procedure . . . . .	40
3.3.	Results . . . . .	42
3.3.1.	Search Times . . . . .	42
3.3.2.	Fixation Distributions . . . . .	44
3.3.3.	Search Target Sizes . . . . .	45
3.3.4.	Search Target Categories . . . . .	46
3.3.5.	Error Rates . . . . .	47
3.4.	Discussion . . . . .	48
3.4.1.	Clutter-Specific Effects . . . . .	49
3.4.2.	Characterizing Set Size . . . . .	50
3.4.3.	Target-Specific Effects . . . . .	54
3.5.	Conclusion . . . . .	55
<b>4.</b>	<b>Study II . . . . .</b>	<b>56</b>
4.1.	Introduction . . . . .	56
4.2.	Methods . . . . .	59
4.2.1.	The Exemplar Clutter Metric . . . . .	60
4.2.1.1.	Representing the Search Target . . . . .	60
4.2.1.2.	Representing the Search Background . . . . .	60

4.2.1.3.	Measuring Similarity . . . . .	61
4.2.2.	The Category Clutter Metric . . . . .	61
4.2.2.1.	Representing the Categorical Search Target . . .	62
4.2.2.2.	Representing the Search Background . . . . .	62
4.2.2.3.	Measuring Similarity . . . . .	63
4.2.3.	The Target-Agnostic Metric . . . . .	63
4.3.	Results . . . . .	64
4.3.1.	The Exemplar Clutter Metric . . . . .	64
4.3.2.	The Category Clutter Metric . . . . .	65
4.3.2.1.	Target Present Images . . . . .	66
4.3.2.2.	Target Absent Images . . . . .	66
4.3.3.	Contributions of Clutter Metrics . . . . .	67
4.4.	Discussion . . . . .	68
4.5.	Conclusion . . . . .	71
<b>5.</b>	<b>General Conclusions . . . . .</b>	<b>72</b>
	<b>References . . . . .</b>	<b>76</b>

## List of Tables

3.1. Pearson correlation coefficients among clutter measures . . . . .	40
3.2. Error rates across conditions . . . . .	48
4.1. Parameter estimates, AICs, and pseudo $R^2$ s for target present images	65
4.2. Parameter estimates, AICs, and pseudo $R^2$ s for target absent images	67
4.3. Parameter estimates, AICs, and pseudo $R^2$ s in target present im- ages for nested models . . . . .	68
4.4. Parameter estimates, AICs, and pseudo $R^2$ s in target absent images for nested models . . . . .	68

## List of Figures

3.1. Example stimulus displays . . . . .	33
3.2. Distribution of clutter values and target size for each category . .	36
3.3. Search task sequence for a trial . . . . .	37
3.4. Example segmented images . . . . .	38
3.5. Average search times as a function of relevant set size . . . . .	44
3.6. Fixation distributions . . . . .	45
3.7. Search times as a function of target size . . . . .	46
3.8. Average search times for different targets in common images . . .	47
3.9. Search times as a function of the object-cue count . . . . .	52
3.10. Distribution of target locations . . . . .	53
4.1. Search times as a function of the exemplar clutter metric and the target-agnostic metric in target present images . . . . .	65
4.2. Search times as a function of the category clutter metric in target present images . . . . .	66
4.3. Search times as a function of the category clutter metric and the target-agnostic metric in target absent images . . . . .	67

# Chapter 1

## Introduction and Objectives

### 1.1 Introduction

Visual search, scanning the visual environment to find an item, is a part of everyday tasks. We constantly engage in visual searches as we complete daily actions such as finding a set of keys in a bag or locating a cellphone in a living room. There are also more crucial domains which require visual search tasks at a professional level, such as radiology. Therefore, understanding the mechanisms of visual search has significant implications. Performance in visual search tasks could depend on a number of factors (see Eckstein, 2011, for a review), including the amount of visual clutter within the environment.

Visual clutter is usually defined as “the state in which excess items, or their representation or organization, lead to a degradation of performance at some task” (Rosenholtz, Li, Mansfield, & Jin, 2005, p. 761). Visual clutter could affect the visual search performance in a number of ways. First, clutter could increase the likelihood of obscuring targets. In natural scenes, objects appear in various positions and orientations, and they can occlude one another. As a scene gets more cluttered, the possibility of occlusions could increase. As a result, visual search in highly cluttered scenes would become more difficult (Bravo & Farid, 2008; Rosenholtz, Li, & Nakano, 2007).

Second, clutter could affect search by increasing the number of possible locations where targets can appear. There a number of studies which looked at the

effect of clutter on search using various type of naturalistic stimuli (e.g., Bravo & Farid, 2004, 2008; Henderson, Chanceaux, & Smith, 2009; Lohrenz, Trafton, Beck, & Gendron, 2009; Neider & Zelinsky, 2011; Rosenholtz et al., 2007). According to these studies, as a scene becomes cluttered, the number of possible target locations also increases. In fact, in natural scenes, clutter is usually used as a stand-in for set size (Rosenholtz et al., 2007). These studies show that performance decreases a function of clutter in natural scenes.

Third, clutter could also affect visual search performance by decreasing the ability to exclude irrelevant information. To investigate this, in a recent study, we designed an experimental technique that modulates the distribution of clutter in synthetic noise displays (Semizer & Michel, 2017). For this study, we defined the relevant clutter in terms of the overlap between the spatial frequency band of the target, a sine-wave grating, and of the background,  $1/f$  noise. In the cluttered condition, the display was tiled uniformly with relevant feature clutter. In the uncluttered condition, we removed the clutter at the irrelevant locations and kept it only at the possible target locations. Additionally, we developed a constrained ideal searcher model. This model was successful in accounting for human performance in these displays, suggesting that clutter impairs performance by forcing observers to consider irrelevant locations. By expanding upon these findings, one of the goals of this dissertation is to investigate whether this effect generalizes to search in natural images.

Defining what clutter is, quantifying the amount of visual clutter in natural images, and using this measure to predict search performance is a difficult task. A reliable measure of clutter is necessary to be able quantify and predict performance in naturalistic tasks. The most known computational models of visual clutter in psychophysics include the following:

- Feature Congestion metric (Rosenholtz et al., 2005) defines clutter in terms

of the saliency of features in a display. A display is considered highly cluttered if it is difficult to draw attention to a newly added item. Clutter is quantified as the amount of local variability in features such as luminance contrast, color or orientation with the display.

- Subband Entropy metric (Rosenholtz et al., 2007) defines clutter in terms of the amount of “disorganization” in a display. If the display contains a lot of redundant information, it would be considered highly cluttered. Clutter is related to the number of bins required to code an image with a given quality, which can be quantified as the entropy within image subbands.
- Edge Density metric (Mack & Oliva, 2004) defines clutter in terms of the density of edges within a display. A display with a lot of edges would be considered highly cluttered. Clutter is quantified by extracting edges from an image and dividing it by the image size.
- Segmentation based clutter metric (Bravo & Farid, 2008), which is based on a segmentation algorithm (Felzenszwalb & Huttenlocher, 2004) defines clutter in terms of how the number of regions within a display changes as the scale of regions does. Clutter is quantified by segmenting images at multi-scales, fitting a power-law to the relationship between the number of segments and the scales. The constant of proportionality is used as the measure of clutter.
- Proto-object model (Yu, Samaras, & Zelinsky, 2014) defines clutter in terms of the number of proto-objects in a display. Proto-objects are defined as “similar” segments within a display and formed by superpixels which share similar features using segmentation and clustering.

Although there can be many different ways to improve the current measures of visual clutter, we are specifically interested in the fact that these measures do

not take the properties of search targets into account. However, as we have shown in our previous study (Semizer & Michel, 2017), performance differs depending on the similarity between features of search targets and features of search background. In particular, when the search target and the background share similar features such as spatial frequency content, search becomes more difficult (Semizer & Michel, 2017).

Unlike aforementioned models, there are other models of clutter from the field of image optics, which quantify clutter in terms of target-background similarity. Some of these measures are purely mathematical while others incorporate properties of the visual system into their calculations. Most common measures in engineering include the following:

- The statistical variance (SV, Schmieder & Weathersby, 1983) and Silk’s statistical variance (SSV, Silk, 1995) metrics define clutter in terms of the variance of intensities at the adjoining regions of a display. However, this metric is too simple to apply to complex images and the only feature of target that is taken into account is its size.
- The probability of edge (POE, Tidhar, Reiter, Avital, & Hadar, 1994) metric defines clutter with the assumption that human visual system is sensitive to edges in a display. However, the only feature of a target that is taken into account is its size. Additionally, this metric requires a user defined threshold.
- The target structure similarity (TSSIM, Chang & Zhang, 2006) metric defines clutter in terms of the similarity between the features of a target and a background, such as intensity, contrast, and structure. However, its performance depends on user defined constants.
- The contrast sensitivity function (CSF, Chu, Yang, & Qian, 2012) based

metric defines clutter in terms of the similarity between a target and a background in the frequency domain after filtering out the non-visible frequencies to human eye and adjusting remaining frequencies with appropriate weights. However, it is subject to shift-variance.

- The masked target transform volume (MTTV, Moore, Camp, Moyer, & Halford, 2010a, 2010c) metric defines clutter in terms of the shared spatial frequencies between a target and a background. This metric assumes that the frequency overlap between the background and the target impedes search performance. It attempts to account for the human visual system by using a contrast sensitivity function and a foveal mask to sub-window the image.

Although these metrics measure clutter in terms of target-background similarity, they focus only on a particular instance of a target in a given image rather than a categorical target. As a result, these metrics cannot measure the effect of clutter on search performance for categorical targets. Additionally, these metrics cannot make predictions for search in target absent images. Moreover, although some of these metrics attempt to account for human visual system (e.g., Chu’s CSF or MTTV metrics), they do not explain the effect of similarity on clutter in the periphery. This is an especially important point because clutter has a greater effect on search in the periphery rather than in the fovea.

Unlike previous metrics, we define two types of target-background similarity: exemplar similarity and category similarity. Exemplar similarity refers to the overlap between the features of the actual target that is present in an image (at the exemplar level) and the features of the actual background. Previous clutter metrics focus on the exemplar similarity while ignoring the category similarity, which refers to the overlap between the representation of the features of the target

(at the category level) and the features of the actual background. Exemplar similarity can simply be measured by comparing the features of an instance of the target and the background in an image. Category similarity can be measured by calculating the distribution of features of targets from a given category across a number of images and comparing it to the features of the actual background. One of the goals of this dissertation is to introduce a new method of quantifying clutter based on these two aspects of target-background similarity in order to predict the effect of clutter on search performance.

For example, think about an observer who is searching for a black squirrel on a ground covered with brown leaves. Since the squirrel does not share a lot of features with the background, the exemplar similarity would be low. Given that most squirrels are brown; the distribution of squirrel colors would likely have a sharp peak at the color brown. Since a “typical” squirrel shares features with the background, the category similarity would be high. This scenario is particularly interesting to separate the effects of exemplar and category similarity on quantifying clutter. Additionally, the mere effect of category similarity on clutter would be observed in a case where the target is absent in a scene.

Now, assume that the actual squirrel on the leaves is brown. It would easily blend with the background, and both types of similarity would be high. Similarly, if the brown squirrel was on the grass, rather than on the brown leaves, both types of similarity would be low. Clutter would have the most impact on performance when both types of similarity are high and the least impact on performance when both types of similarity are low.

The features of a categorical search target and its background can be measured within natural images. Exemplar similarity can be quantified by measuring the overlap of these features between the target and the background. The distribution of features for a given target category can be used to measure category

similarity. In target absent cases, the category level target features (e.g., average of the distribution) can be used to represent features of a “typical” target. This dissertation attempts quantifying these similarity metrics.

## 1.2 Objectives

The aim of this dissertation is to investigate how clutter affects search for categorical targets by answering the following questions:

1. What is the effect of clutter on search performance for categorical targets when the set size (i.e., the number of possible target locations) is controlled?
2. What visual features of categorical search targets and backgrounds are important in measuring clutter?

### 1.2.1 Study I: What is the effect of clutter on search performance for categorical targets when the set size is controlled?

In this study, we controlled for the number of possible target locations by fixing the relevant set size. When the set size is fixed, clutter would have an effect on search performance by forcing observers to consider irrelevant locations. Our stimuli in this study was a set of natural images with minimal contextual effects. This ensured that search targets were likely to appear anywhere in an image. Most frequent objects were chosen as categorical search targets (cellphones, keys, etc.), and their locations and sizes were measured. The amount of clutter in each image were quantified using a version of an existing measure of clutter (Bravo & Farid, 2008) which was based on a method of image segmentation (Felzenszwalb

& Huttenlocher, 2004). Human overt search performance were investigated using these stimuli.

### **1.2.2 Study II: What visual features of categorical search targets and backgrounds are important in measuring clutter?**

The target-background similarity is an important factor in predicting search performance in natural images. In this study, we defined two new clutter metrics (i.e., the exemplar clutter metric and the category clutter metric) which account for target-background similarity. Then, we applied these metrics to a set of natural images and measured existing clutter in these images. Additionally, performance of these metrics were compared to performance of an existing clutter metric to investigate how well they predict human search performance in these images.

## Chapter 2

# Visual Search and Clutter

### 2.1 Visual Search

Visual search is used daily both in mundane tasks such as searching for keys in a container or for a friend in a crowd and in high-stakes tasks such as searching for a tumor in an X-ray image or screening for explosives in a baggage. Since visual search tasks are highly prevalent and can have critical consequences, performing these tasks efficiently is of utmost importance. Several factors which mediate the search performance have been identified (for a review, see Eckstein, 2011). Some of these factors limit the performance while others improve it.

#### 2.1.1 What Limits Search?

Perceptual and cognitive limitations on visual search include information loss in the periphery, limited memory capacity and limited attentional processes. In this section, I will describe the ways in which these three factors limit search by providing empirical evidence from the field.

##### 2.1.1.1 Information Loss in the Periphery

The human visual system has its highest acuity and spatial resolution at the center of the retina, i.e., at the fovea (De Valois & De Valois, 1988; Levi, Klein, & Aitsebaomo, 1985). Information starts to get lost as moved away from the fovea, or as visual eccentricity increases, in the periphery. The sensitivity falloff

in the peripheral vision compared to the foveal vision results from a number of biological constraints. Firstly, there are less photoreceptors with larger receptive fields in the periphery. Secondly, a larger number of photoreceptors gives rise to a single ganglion cell in the periphery. Lastly, the fovea is favorably represented at the visual cortex (De Valois & De Valois, 1988; R. O. Duncan & Boynton, 2003; Fischer, 1973).

It is well established that as visual eccentricity increases, visual search performance suffers, which is illustrated by lower accuracy or longer search times (Carrasco, Evert, Chang, & Katz, 1995; Geisler & Chou, 1995). However, the resolution degradation in the periphery is not the only factor that can account for the performance drops in detecting and localizing peripheral targets. Another factor which limits peripheral vision is its vulnerability to clutter. Clutter influences visual search by reducing accuracy and increasing search times; and it gives rise to crowding, where the ability to recognize items in the periphery is impaired by neighboring features (Bouma, 1970; Levi, 2008; Pelli, Palomares, & Majaj, 2004; Rosenholtz et al., 2007). Intrinsic position uncertainty has been proposed as one of the mechanisms underlying the effect of clutter and crowding. Intrinsic uncertainty refers to the uncertainty regarding the source of a perceived stimulus and it limits visual search as a function of visual eccentricity (Michel & Geisler, 2011). Although there is not a consensus on how to define clutter, several computational models have been developed to quantify clutter in naturalistic scenes (e.g., Bravo & Farid, 2008; Neider & Zelinsky, 2011; Rosenholtz et al., 2005, 2007; Yu et al., 2014).

There are ways to measure the information loss in the periphery. Visibility maps are used to measure observers' sensitivity using psychophysical methods (Najemnik & Geisler, 2005). Many computational models of visual search take observers' visual sensitivity into account (e.g., Najemnik & Geisler, 2005; Zelinsky,

2008). Despite its limitations, peripheral vision provides observers with useful information in visual search tasks. Peripheral information can influence where observers will shift their gaze next during search.

### **2.1.1.2 Limited Memory Capacity**

Our memory capacity is limited (Brady, Konkle, & Alvarez, 2011; Luck & Vogel, 1997; Miller, 1956). Earlier studies on memory capacity investigated the number of items memory can simultaneously hold (e.g., Luck & Vogel, 1997) while more recent studies focused on the nature of memory representations (e.g., Brady et al., 2011). It is intuitive to expect that limited memory capacity constraints visual search because observers need to remember the target features, and they need to avoid revisiting the locations they already examined to search effectively. However, whether visual memory plays a role in visual search is still open to discussion (Hollingworth, 2006; Woodman & Chun, 2006). Although most studies suggest that memory does not seem to limit search or scene representations (e.g., Horowitz & Wolfe, 1998), there is still evidence that memory guides search (e.g., Kristjánsson, 2000). For example, observers tend to fixate less on previously visited locations while searching, suggesting that memory may play a role in visual search (Klein, 1988; Klein & MacInnes, 1999).

### **2.1.1.3 Limited Attentional Processing**

Attention helps us to filter out irrelevant information (i.e., noise). As a result, we can process what is important for the purposes of the task at hand (i.e., signal) more efficiently. However, attentional resources are limited; we can only attend to a limited number of things at a time, either by dividing our attention among stimuli or by selectively attending to particular set of stimuli. Limited attentional processing is one of the constraints on visual search.

Previous research has revealed three types of attention that are relevant for visual search: spatial attention, feature-based attention, and object-based attention (Carrasco, 2011). Spatial attention refers to directing attention to specific locations either overtly (with eye movements) or covertly (without eye movements). Feature-based attention refers to covertly deploying attention to particular features while object-based attention refers to directing attention to particular objects. These attentional processes can guide visual search in various ways, such as by directing eyes to certain spatial locations or to certain target features (e.g., color or shape of the target). Additionally, attentional pre-cuing has been shown to enhance performance while detecting and discriminating targets or while searching for targets (Baldassi & Burr, 2000; Carrasco, Giordano, & McElree, 2004; Carrasco & Yeshurun, 1998). Although covert attention improves visual search performance, the mechanisms through which this occurs is not clear. Some of the proposed processes include: (a) attention enhances processing of the relevant information (Cameron, Tai, & Carrasco, 2002; Lu & Doshier, 1998, 2000), (b) attention helps to filter out irrelevant noise (Doshier & Lu, 2000a, 2000b), and (c) attention reduces spatial uncertainty (Eckstein, Thomas, Palmer, & Shimozaki, 2000; Verghese, 2001). The effect of attention on visual search might result from a combination of these processes (Carrasco, 2011). However, because the mechanisms through which attention guides search is not well-defined, it is challenging to quantify the effect of attention on visual search.

### **2.1.2 What Improves Search?**

Although the aforementioned factors limit search, there are other mechanisms through which performance can be improved. For example, perceptual learning occurs through practice. Training improves perceptual and cognitive processes, minimizes uncertainty about the location of targets, and improves performance

in visual search. With training, observers develop a deeper understanding of the properties of targets (e.g., their shape or size), possible locations of targets within the search environment, or the frequency of encountering targets. This prior information in turn guides observers' search behavior. In this section, I will describe how prior knowledge about the target properties, search environment, and target prevalence influences visual search.

### **2.1.2.1 Target Properties**

To effectively perform a visual search task, observers need to have a representation of what they are looking for. Observers use information regarding relevant features of the target to guide their search behavior. For example, while trying to locate a white car in the parking lot, the observer may search for specifically white targets. Prior information about target properties guides selection of eye movements and perceptual judgments (Eckstein, Beutter, Pham, Shimozaki, & Stone, 2007; Findlay, 1997; Murray, Beutter, Eckstein, & Stone, 2003; Rajasekar, Bovik, & Cormack, 2006; Williams, 1967). However, because of the environment variability, such as occlusions or lighting conditions, actual appearance of target features might dramatically deviate from observers' prior expectations or knowledge. Similarly, observers might have an inherent uncertainty regarding the actual properties of the target (e.g., its shape or size). As a result, they might search for a longer period of time and possibly leading to the failure to detect the target.

Most psychophysical studies do not assume uncertainty regarding the target parameters and they use the signal-known-exactly (SKE) paradigm in which observers know exactly what the target looks like (Brettel, Berry, & Smith, 2007; Burgess, 1994). However, in real practice observers do not always have perfect knowledge about the properties of the targets. Other more naturalistic methods

were also developed such as signal-known-statistically (SKS) or signal-known-exactly but variable (SKEV) paradigms (Castella et al., 2009; Y. Zhang, Pham, & Eckstein, 2005), where observers have the knowledge about the distribution of target features. Although, SKS and SKEV methods have higher external validity, they are not frequently used because they are more challenging to model. Regardless of the specific method used, uncertainty about the target properties impairs performance while detecting targets (Burgess & Ghandeharian, 1984a; Cohn & Lasley, 1974; Cohn & Wardlaw, 1985; Eckstein & Whiting, 1996). Additionally, there is strong evidence that uncertainty about the target properties impairs search performance (e.g., Bravo & Farid, 2009; Vickery, King, & Jiang, 2005).

Low levels of uncertainty regarding the properties of search targets can improve the performance in search tasks. For example, experts engaging in visual search report prior information about the target parameters as one of the major factors in a successful search (Eckstein, 2011). In radiology, experts’ deeper knowledge regarding properties of perturbations enables them to mark potential areas of abnormalities faster than novices, leading to a superior performance in target detection (Drew, Evans, Vö, Jacobson, & Wolfe, 2013; Lesgold et al., 1988; Reingold & Sheridan, 2011). Over years of practice, experts develop templates for normal and abnormal structures. Additionally, they develop a better understanding of the anatomical structure. They learn about the typical locations of abnormalities so they compare those locations to the “normal tissue” template that they have in their minds (Donovan & Litchfield, 2013).

### **2.1.2.2 Search Environment**

Search environment also guides visual search. For example, while searching for a painting, observers would direct their gaze to walls. Their knowledge of

where paintings are frequently placed decreases the number of locations they need to search. In this sense, scene context reduces the amount of uncertainty regarding the possible target locations (i.e., extrinsic uncertainty), and improves search performance.

Scene context has been shown to guide eye movements to more probable target locations during visual search tasks (Castelhano & Heaven, 2011; Henderson, Weeks, & Hollingworth, 1999; Neider & Zelinsky, 2006; Oliva & Torralba, 2006; Torralba, Oliva, Castelhano, & Henderson, 2006). Scene guidance benefits from humans’ ability to extract the “gist” of scenes very quickly (Friedman, 1979; Potter, 1976; Potter & Levy, 1969; Schyns & Oliva, 1994). Humans can rapidly categorize real world scenes as urban or natural (Greene & Oliva, 2009), or as indoor or outdoor (Fei-Fei, Iyer, Koch, & Perona, 2007).

While prior information about the search environment guides search, the uncertainty about the environment impairs search performance. If distractors have shared parameters with targets, errors are more probable and search would take a longer time. Luckily, human observers seem to adapt to constraints for the search environment to some extent. While searching in a noisy environment, observers seem to adjust their search strategies or target templates to overcome the environment’s uncertainty (Burgess, Li, & Abbey, 1997; Y. Zhang, Abbey, & Eckstein, 2006).

### **2.1.2.3 Target Prevalence**

Prior information regarding the prevalence of targets guides search behavior. If observers know the probability of the targets’ occurrences, they then can adjust their responses accordingly to optimize the performance. For example, if target prevalence is low, reporting a target as absent would maximize performance because in most cases the target would be absent. Similarly, if target prevalence

is high, the best strategy is to say “target-present” in most of the cases. Low prevalence typically leads to a more *conservative* decision criterion (i.e., more willingness to say target is absent) while high prevalence typically leads to a more *liberal* decision criterion (i.e., more willingness to say target is present). Using signal detection theory, the optimal decision criterion can be calculated for a particular level of target prevalence (Green & Swets, 1966).

There is strong evidence suggesting that low target prevalence leads to a higher rate of false-negative errors and early termination of search (e.g., Fleck & Mitroff, 2007; Kunar, Rich, & Wolfe, 2010; Rich et al., 2008; Wolfe, Horowitz, & Kenner, 2005; Wolfe et al., 2007), even in well-trained observers (Wolfe, Brunelli, Rubinstein, & Horowitz, 2013). The mechanisms behind the higher rate of false-negatives in low prevalence conditions are explained through (a) a shift in decision criterion without a decrease in sensitivity (Green & Swets, 1966; Wolfe et al., 2007), (b) motor-response errors (Fleck & Mitroff, 2007), or (c) early termination of search (Wolfe & Van Wert, 2010). Additionally, the task complexity is a factor in determining the source of errors in low prevalence conditions (Rich et al., 2008).

## 2.2 Visual Clutter

Although everyone intuitively knows what clutter is, clutter is an ill-defined concept that influences performance in a wide range of professions such as airport security agents searching for prohibited items in luggages or pilots monitoring flight decks. The harmful effects of clutter on performance are documented in several areas of research such as psychology, engineering, computer science and marketing, but there is still no agreement on how to define or measure clutter. In this section, I will discuss the proposed definitions of clutter and the current techniques in quantifying clutter, in the area of visual search in particular.

### 2.2.1 How to Define Clutter?

How would you go about defining clutter? Is it the number of objects in a scene? If so, how would you define an object? Or, is it the organization (or disorganization) of items in a scene? Is the similarity of background features to target features important in defining clutter? If so, how would you characterize similarity? These questions are challenging and the definition of clutter depends on the area of research.

According to the most common definition, clutter is related to the number of objects in a display. This approach to clutter is similar to traditional set-size effects (Palmer, 1994, 1995), which state that search time increases as the number of distractors increases (Wolfe, 1994). Similarly, based on this definition, in natural scenes, the density of items in a display correlates with the amount of clutter in the display (van den Berg, Cornelissen, & Roerdink, 2009). Although this definition of clutter is intuitive, it may not be easy, or even possible, to enumerate the number of items or to estimate the density of items in natural scenes. Additionally, in natural scenes, contextual information effectively reduces the set size (Castelhano & Heaven, 2011; Neider & Zelinsky, 2006; Oliva & Torralba, 2006; Torralba et al., 2006).

Another common definition of clutter focuses on the layout of items in a display. This account considers the arrangement or structure of items in a scene in addition to their density (Bravo & Farid, 2008; Rosenholtz et al., 2005; van den Berg et al., 2009). For example, high variability of luminance, contrast, orientation, and color in a display (Rosenholtz et al., 2005), high density of edges in a scene (Mack & Oliva, 2004), or large number of regions in a display (Bravo & Farid, 2008) are associated with clutter. Eliminating variability of such features (e.g., losing color information) might help to decrease effects of clutter but in turn

might increase the task difficulty.

Clutter can also be defined in terms of the target-background (or target-distractor) similarity in a display. The similarity between features of search targets and features of search background has already been shown to impair performance in search tasks (J. Duncan & Humphrey, 1989; Wolfe, Oliva, Horowitz, Butcher, & Bompas, 2002). There are other measures of clutter in the field of image optics, which quantify clutter using features within a scene that are similar to features of a search target (Camp, Moyer, & Moore, 2010; Chu et al., 2012; He, Zhang, Liu, & Chang, 2008). However, these measures consider only a limited range of features such as target size.

The aforementioned definitions view clutter primarily as a scene property. Another common definition evaluates clutter using task performance while taking different aspects of clutter into account. According to this perspective, clutter is defined as “the state in which excess items, or their representation or organization, lead to a degradation of performance at some task” (Rosenholtz et al., 2005, p. 761). This account also links clutter to density or structure of items in a display.

The choice of how to define clutter somewhat determines how to measure clutter in a display. There are several measures of clutter which takes different approaches in quantifying existing clutter in natural images to predict search performance. Next, I will discuss the most common of these methods.

### **2.2.2 How to Measure Clutter?**

Clutter is measured using various approaches, such as developing objective clutter metrics, directly manipulating clutter in a display and measuring task performance, or collecting subjective judgments by asking observers to rank displays based perceived clutter. Here I will primarily focus on objective clutter metrics that have been used in two distinct areas of research, psychophysics and image

optics, to examine performance in search tasks. The measures in psychophysics can predict search performance as a function of clutter but they do not account for the similarity between features of search targets and features of search background. The measures in image optics account for some target features but they are limited and mostly not for categorical targets.

### **2.2.2.1 Clutter Measures in Psychophysics**

The existing clutter measures in visual psychophysics have shown that as the amount clutter in a scene increases, search performance decreases (Bravo & Farid, 2008; Henderson et al., 2009; Mack & Oliva, 2004; Rosenholtz et al., 2007; Neider & Zelinsky, 2011). I will describe the most common of these clutter measures below.

#### **2.2.2.1.1 The Edge Density Metric**

The Edge Density (Mack & Oliva, 2004) metric quantifies clutter as a function of the number of edges in a display, with an assumption that more edgy displays would be more cluttered. Clutter is estimated by first filtering the image to extract the edge information and then calculating the density of edges. The Edge Density metric of clutter has been shown to predict visual search performance despite its simplicity (e.g., Henderson et al., 2009; Neider & Zelinsky, 2011; Rosenholtz et al., 2007).

#### **2.2.2.1.2 The Feature Congestion Metric**

The Feature Congestion (Rosenholtz et al., 2005, 2007) metric computes clutter as a function of the local variability in features such as color, orientation, and luminance contrast within an image. As the variability of features in an image increases, so does its level of clutter. This metric is based on the Statistical Saliency

Model (Rosenholtz, 1999; Rosenholtz et al., 2007), which quantifies saliency of features using their local variability. The saliency is computed as

$$\Delta = \sqrt{(T - \mu_D)' \Sigma_D^{-1} (T - \mu_D)} \quad (2.1)$$

where  $T$  is a vector representing a particular feature, and  $\mu_D$  and  $\Sigma_D$  are the mean and the covariance of a set of feature vectors, respectively. As the saliency of a target increases, search performance is expected to increase as well.

The Feature Congestion metric is implemented by creating a Gaussian pyramid of an image at multiple scales, computing local variability of features (color, orientation, and luminance contrast) at each scale, combining variability across scales and features, and averaging over the entire image. This metric is one of most commonly used measures of clutter in visual search studies (e.g., Henderson et al., 2009; Neider & Zelinsky, 2011; Rosenholtz et al., 2005, 2007).

#### 2.2.2.1.3 The Subband Entropy Metric

The Subband Entropy (Rosenholtz et al., 2007) metric quantifies clutter in relation to the number of bits necessary for encoding an image. It is based on the idea that if an image has a lot of redundancies, then it would be easily encoded requiring less number of bits; therefore it would be considered less cluttered. The Subband Entropy metric is computed by decomposing an image into wavelet subbands using Steerable Pyramids (Simoncelli & Freeman, 1995), binning coefficients within each subband, and computing entropy using the Shannon's formula

$$H = \sum_i -p_i \log(p_i) \quad (2.2)$$

where  $p_i$  is the probability distribution of binned wavelet coefficients in bin  $i$ . Finally, clutter is estimated by summing entropies across subbands and channels

while imposing certain weights.

#### **2.2.2.1.4 The Segmentation Based Clutter Metric**

The segmentation based clutter metric (Bravo & Farid, 2008) quantifies clutter using a segmentation algorithm (Felzenszwalb & Huttenlocher, 2004) which counts the number of regions in an image. This metric is computed as follows: First, images are segmented at a range of different scales using the segmentation algorithm. The relationship between the number of segments and the scale of segmentation can be described using a power-law given by

$$y = \alpha k^\beta \quad (2.3)$$

where  $y$  is the number of segments,  $k$  is the scale of segmentation. These fits yield two parameters, an exponent ( $\beta$ ), and a constant of proportionality ( $\alpha$ ). The fitted exponent is more or less constant across images, whereas the constant of proportionality varies dramatically across images. This constant of proportionality is used as the measure of clutter.

#### **2.2.2.1.5 The Proto-Object Model**

The proto-object model (Yu et al., 2014) quantifies clutter as the number of regions composed of coherent features (i.e., proto-objects) in an image. This model is implemented as follows: First, an image is segmented into regions that share similar color features (i.e., superpixels). Then, neighboring regions which share similar color features are clustered together to form proto-objects. The number of proto-objects is used as the measure of clutter. The proto-object model has been shown to correlate well with behavioral ranking of a set of images based on clutter.

### 2.2.2.2 Clutter Measures in Image Optics

The existing clutter measures in the area of image optics typically quantify clutter based on the similarity between the target and the background. Similarity is measured using different techniques, some of which are purely mathematical while others attempt to represent properties of the human visual system. I will describe the most common clutter measures in optical imaging below.

#### 2.2.2.2.1 The Statistical Variance Metric

The statistical variance (SV) metric (Schmieder & Weathersby, 1983) quantifies clutter as a function of the luminance variance within an image. First, the image is divided into a grid composed of cells. Each cell has a size twice the size of the target in any dimension. Then, clutter is estimated by computing the root mean square of grayscale variance of these cells using the formula

$$SV = \sqrt{\frac{1}{N} \sum_{i=1}^N \sigma_i^2}, \quad (2.4)$$

where  $N$  is the number of cells and  $\sigma_i^2$  is the variance of cell  $i$ .

Although the calculation of the SV metric is straightforward, it has some disadvantages. For example, this metric uses non-overlapping sections which are subject to shift-variance, which might result in drastically different estimates. Silk (1995) modified this metric so that the shifts in the image do not lead to changes in the measure (i.e., shift-invariance). However, both SV metric and Silk's version are too simple to apply to complex images and the only feature of target that is taken into account is its size.

#### 2.2.2.2.2 The Probability of Edge Detection Metric

The probability of edge detection (POE) metric (Tidhar et al., 1994) estimates clutter as a function of the number of edges that exceeds a chosen threshold in an image. This metric is implemented as follows: First, the image is divided into different sections. Similar to SV metric (Schmieder & Weathersby, 1983), each section has a size twice the size of the target in each dimension. Second, a filtering process is applied to enhance the edges in the image. Finally, the number of edge points above a pre-defined threshold in each section are computed and averaged across sections given by

$$\text{POE} = \sqrt{\frac{1}{N} \sum_{i=1}^N \text{POE}_{i,T}^2}, \quad (2.5)$$

where  $N$  is the number of sections,  $\text{POE}_{i,T}$  is the number of points that exceeds the threshold  $T$  in section  $i$ .

The POE metric is different than the SV metric (Schmieder & Weathersby, 1983) in the sense that it is sensitive to the number of edge points rather than the magnitude of these points and it outperforms the SV metric in accounting for human detection performance (Tidhar et al., 1994). However, the POE metric uses a user-dependent threshold which might lead to unstable estimates. More importantly, both of these metrics consider only the target size as relevant target feature.

#### 2.2.2.2.3 The Target Structure Similarity Metric

The target structure similarity (TSSIM) metric (Chang & Zhang, 2006) quantifies clutter by comparing intensity, statistical variance, and structure between target and background areas in an image. TSSIM metric is implemented in the

following steps: First, the image is divided into several sections. Each section has a size twice the size of the target in each dimension. Second, TSSIM is measured for each section by computing similarity of luminance (mean), contrast (standard deviation) and structure (covariance) between the target and the particular section and multiplying them using some constants. In its simplest form, this can be expressed as

$$\text{TSSIM}(T, B_i) = \frac{4\mu_T\mu_{B_i}\sigma_{TB_i} + C}{(\mu_T^2 + \mu_{B_i}^2)(\sigma_T^2 + \sigma_{B_i}^2) + C}, \quad (2.6)$$

where  $\mu_T$ ,  $\mu_{B_i}$  and  $\sigma_T^2$ ,  $\sigma_{B_i}^2$ , and  $\sigma_{TB_i}$  are mean, variance and covariance of the target section and the background section  $i$ , respectively, while  $C$  is a user-defined constant. Finally, TSSIM for the whole image is computed by taking the mean square root of the TSSIM across sections given by

$$\text{TSSIM}_{rms} = \sqrt{\frac{1}{N} \sum_{i=1}^N \text{TSSIM}(T, B_i)^2}. \quad (2.7)$$

This metric was originally developed to measure clutter in grayscale images and later it has been extended to color images (Chang, Zhang, Liu, Yang, & Li, 2010). In this adaptation (CTSSIM), each color image is first converted to CIELAB color space. Then, TSSIM metric is measured in each dimension and combined using weighted means. Both TSSIM and CTSSIM were reported to correlate well with human detection performance; however, their performance depends on user-defined constants.

#### 2.2.2.2.4 The Contrast Sensitivity Function Based Metric

Contrast sensitivity function (CSF) based metric (Chu et al., 2012) measures clutter in the frequency domain after filtering out the frequency content that is

not perceived by the human visual system. The CSF-based metric is computed as follows: First, the luminance image is divided into several sections. Each section is twice the size of the target in each dimension. Fourier transform is applied to each section followed by a filtering process using a CSF. This was done for each pair of background and target section using the formula

$$MS\_CM_i = \sqrt{\sum_{u,v} |B_i(u,v) - T(u,v)|^2 / (w \times h)}, \quad (2.8)$$

where  $T(u,v)$  and  $B_i(u,v)$  represent filtered Fourier transform of the target section and the background section  $i$  at a particular frequency  $(u,v)$ , respectively; and  $(w \times h)$  is the size of the section. Finally, the  $CSFC_{MS}$  is computed by averaging the  $MS\_CMs$  over the entire image given by

$$CSFC_{MS} = \frac{1}{N} \sum_{i=1}^N MS\_CM_i, \quad (2.9)$$

where  $N$  is the number of sections. One of the disadvantages of CSF-based metric is that it uses frequency content of non-overlapping sections and the magnitude information of target section is prone to edge effects resulting in shift-variance.

#### 2.2.2.2.5 The Masked Target Transform Volume Metric

The masked target transform volume (MTTV) metric (Moore et al., 2010a, 2010c) quantifies clutter in terms of the similarity of the spatial frequency content between the target and the background in image. The MTTV metric is computed as follows: First, the luminance image is divided into subwindows. Each window is weighted with a Gaussian foveal mask. Then, a function to represent contrast sensitivity (Barten, 1990) is applied to each window. The similarity of spatial frequency content of each background window to the target window is computed

using Fourier transforms and then averaged, resulting a single clutter measure for the entire image given by

$$\text{MTTV} = \frac{1}{N} \sum_{i=1}^N \frac{\sum_{f_m, f_n} \text{CSF}(f_m, f_n) \min\{|F[t(m, n)]|, |F[b_i(m, n)f(m, n)]|\}}{\text{CSF}(f_m, f_n) |F[t(m, n)]|}, \quad (2.10)$$

where  $N$  is the number of subwindows, CSF is the contrast sensitivity function (Barten, 1990),  $t(m, n)$  is the target section,  $b(m, n)$  is the background section, and  $f(m, n)$  is the Gaussian foveal mask.

The MTTV metric has some advantages over other metrics. It is shift-invariant and independent of arbitrary thresholds, and it considers some characteristics of human visual system. The original metric was developed for single targets. Later, it was generalized to generic targets (Camp, Moyer, & Moore, 2013). However, the generic targets were created by superimposing single targets, which does not necessarily reflect target properties at categorical level.

## Chapter 3

### Study I

#### 3.1 Introduction

Interacting with the world involves, as frequent and ubiquitous subtasks, the detection and localization of objects in our visual environment. These subtasks are called visual searches. One fundamental property common to all visual searches is uncertainty regarding the positions of target objects. This study examines the properties of the visual environment and of the visual system that contribute to this position uncertainty. In particular, our goal is to investigate how visual clutter affects performance when observers search natural images for categorical targets.

Position uncertainty can be due to either extrinsic or intrinsic sources. For example, an observer searching an unfamiliar bookshelf for a particular book will probably have some uncertainty about the location of the book. In this case, (i.e., when the observer does not know the book's location *a priori*) the uncertainty is a result of imprecise specification of the likely target location. This type of position uncertainty increases with the number of potential target locations and is called *extrinsic* position uncertainty. However, even when the observer is familiar with the bookshelf and knows the order of its books, she might still have a hard time localizing the book in the visual periphery. This uncertainty is a result of the limitations intrinsic to the visual system, such as the limitations of peripheral

vision and visual memory, and it is called *intrinsic* position uncertainty.

Regardless of whether it is extrinsic or intrinsic, position uncertainty impairs performance for detecting, discriminating, and localizing stimuli. This is indicated by decreases in detection and localization accuracy (Burgess & Ghandeharian, 1984b; Eckstein et al., 2000), by increases in detection thresholds (Cohn & Wardlaw, 1985; Palmer, Verghese, & Pavel, 2000), and by increases in search times (Egeth, Atkinson, Gilmore, & Marcus, 1973; Treisman & Gelade, 1980). While research on the effects of position uncertainty has typically focused on extrinsic sources of uncertainty (e.g., Bochud, Abbey, & Eckstein, 2004; Burgess & Ghandeharian, 1984b; Swensson & Judy, 1981), a few studies have explicitly focused on intrinsic sources (e.g., Michel & Geisler, 2011; Pelli, 1985; Tanner, 1961). Evidence from these studies, and from studies of visual crowding (e.g., Bouma, 1970; Levi, 2008; Pelli et al., 2004, 2007) suggests that the ability to identify and localize features declines systematically in the periphery. Indeed, position uncertainty has been repeatedly implicated as a primary contributor to crowding (Krumhansl & Thomas, 1977; Popple & Levi, 2005; Wolford, 1975). For example, similar to crowding (Bouma, 1970; Levi, 2008; Levi, Hariharan, & Klein, 2002), intrinsic position uncertainty also increases approximately linearly with eccentricity (Michel & Geisler, 2011). Moreover, the eccentricity-dependent effects of position uncertainty seem to persist in search tasks requiring eye movements, i.e., overt search tasks (Semizer & Michel, 2017).

As an inherent property of the observer’s visual system, intrinsic position uncertainty cannot be experimentally controlled. However, its effect on performance can be observed by manipulating the visual environment. In a recent study, Semizer and Michel (2017) introduced an experimental technique that modulates the effects of intrinsic uncertainty independently of extrinsic uncertainty by manipulating the distribution of clutter in synthetic noise displays. Using this

technique, the authors showed that intrinsic position uncertainty substantially limits overt search performance and that its effects are especially evident when the amount of extrinsic uncertainty is controlled. Does this result generalize to real-world searches?

In many ways, synthetic visual stimuli have been incredibly useful for vision research. Synthetic stimuli provide researchers with a great deal of flexibility and control, enabling them to manipulate individual stimulus features and to determine how these contribute to performance in a variety of tasks. In visual search, for example, measuring performance in synthetic search displays has allowed researchers to discover how observers use information about peripheral target visibility to select fixations (Geisler, Perry, & Najemnik, 2006; Najemnik & Geisler, 2005, 2008; Michel & Geisler, 2009; Vergheze, 2012; S. Zhang & Eckstein, 2010), how intrinsic position uncertainty and clutter in the periphery degrade performance (Michel & Geisler, 2011; Rosenholtz, Huang, Raj, Balas, & Ilie, 2012; Semizer & Michel, 2017), how the template for known search targets is structured (Eckstein et al., 2007), and how observers integrate information about the target across fixations (Caspi, Beutter, & Eckstein, 2004; Kleene & Michel, 2018), all while controlling extraneous properties of the search display (e.g., spectral spatial frequency statistics, environmental contingencies, target location probabilities, etc.) in ways that would be difficult or impossible with natural scenes. However, their highly controlled nature means that synthetic displays may provide only limited insight into how observers search in naturalistic settings.

For example, the search targets used in synthetic displays typically exhibit very little variability across trials, and observers are therefore assumed to represent them with little uncertainty. In contrast, the targets of natural searches typically exhibit many sources of variability. Objects in natural scenes appear

in various positions and orientations, occlude one another, and change appearance depending on the lighting conditions. Moreover, individual exemplars may vary considerably within a natural object category. These sources of variability introduce additional uncertainty that might overwhelm any effects of intrinsic uncertainty on search performance. Thus, it is important to verify that the factors that explain search performance in synthetic displays generalize to account for searches in more naturalistic displays.

One of the major challenges associated with naturalistic tasks in the context of visual search is to quantify the amount of clutter in natural images. Unlike in artificial displays, clutter cannot be directly manipulated in natural images. However, a variety of models have been proposed to quantify scene clutter. These include edge density (Mack & Oliva, 2004), feature congestion (Rosenholtz et al., 2005, 2007), subband entropy (Rosenholtz et al., 2007), the scale invariant clutter measure (Bravo & Farid, 2008), the color-clustering clutter (C3) model (Lohrenz et al., 2009), and the proto-object model (Yu et al., 2014). Using these measures, several studies have shown that clutter degrades performance for search in various types of naturalistic displays including geographic maps (Rosenholtz et al., 2007; Lohrenz et al., 2009), quasi-realistic scenes (Neider & Zelinsky, 2011), natural scenes (Henderson et al., 2009, but also see Asher, Tolhurst, Troscianko, & Gilchrist, 2013), images displaying contents of bags (Bravo & Farid, 2008), and photo-collages of objects (Bravo & Farid, 2004, 2008).

However, these findings confound different potential sources of position uncertainty. As a scene becomes increasingly cluttered, the number of possible target locations (i.e., set size) also increases. This increase in set size augments the position uncertainty due to extrinsic sources. At the same time, due to intrinsic sources of position uncertainty, the ability to exclude irrelevant signals in the periphery decreases in highly cluttered scenes (Michel & Geisler, 2011; Semizer

& Michel, 2017). These two concurrent effects of clutter make it challenging to separate the contributions of extrinsic versus intrinsic uncertainty on performance in highly cluttered images.

The goal of the current study was to separate out the contributions of intrinsic versus extrinsic sources of position uncertainty and to characterize them in a naturalistic search task that requires searching natural images for categorical targets. We approached this goal by controlling and manipulating set size independently of clutter, as in Semizer and Michel (2017). Instead of imposing synthetic clutter, we used an existing clutter measure (Bravo & Farid, 2008), chosen for its efficiency and its demonstrated correlation with search performance, to quantify the existing clutter in a set of natural images. The images were sorted into high and low clutter conditions based on this clutter measure. The “relevant set size” (Palmer, 1994, 1995), which governed the extrinsic position uncertainty was varied independently by manipulating the number and positions of cues indicating potential target locations. Within each fixed set size condition, search times increased as a function of increasing clutter, suggesting that clutter degrades overt search performance independently of set size.

## 3.2 Methods

### 3.2.1 Observers

A total of twenty-five observers participated in the study. One of the observers was an author; the remaining observers were naïve to the purpose of the experiment and received compensation for their participation. All observers had normal or corrected-to-normal vision.

### 3.2.2 Apparatus

Stimuli were presented on a 22-in Philips 202P4 CRT monitor at 100 Hz. The resolution was set to  $1280 \times 1024$  pixels. Observers were seated 70 cm away from the display so that the display subtended  $15.8^\circ \times 21.1^\circ$  of visual angle. The stimuli displays were programmed using MATLAB software (Mathworks) and the Psychophysics Toolbox extensions (Brainard, 1997). Observers' eye movement signals were monitored and recorded using an Eyelink 1000 infrared eye tracker (SR Research, Kanata, Ontario, Canada) at 1000 Hz. Head position was stabilized using a forehead and chin rest.

### 3.2.3 Stimuli

Images of natural scenes often contain contextual information that effectively reduces the search set size (Castelhano & Heaven, 2011; Neider & Zelinsky, 2006; Oliva & Torralba, 2006; Torralba et al., 2006). To minimize this contextual information, we chose a set of images displaying the contents of bags in arbitrary arrangements (see Figure 3.1). These images were retrieved from the “What’s in your bag?” group on Flickr<sup>1</sup>. We selected five of the most common objects in the image set (cellphones, glasses, iPods, keys, and pens/pencils) to serve as the categorical search targets. If a target object was present in the image, it was either present as a single instance or, in the case of collective objects, as a single group of instances in close proximity (e.g., keys attached to a keychain). There was never more than one instance or group of the target object present in the image.

---

<sup>1</sup>A subset of images from this group were also used in a search task by Bravo and Farid (2008).



Figure 3.1: Example displays for the low clutter (on the left) and the high clutter (on the right) conditions with keys as the search target. Keys are located near the center in both images. Images are retrieved from “What’s in your bag?” group on <https://www.flickr.com>.

### 3.2.3.1 Creating the Image Data Set

The image data set was created by processing raw images in four separate stages: initial filtering, transformation, labelling, and selection. Each stage was described in detail next.

#### 3.2.3.1.1 Initial Filtering Stage

Images were downloaded and subsequently checked for duplicates and quality (e.g., blurs, artifacts, etc.). We avoided scaling the size of small images up to preserve image quality. Therefore, images whose maximum dimension smaller than the height of the stimulus window (1024 pixels) were excluded.

#### 3.2.3.1.2 Transformation Stage

The clutter measure used in our experiment is sensitive to the image size (see the Measuring Clutter section). To control for any potential effects of image size on quantifying clutter, we resized the minimum dimension to 1024 pixels.

Next, we considered the variability in color across images. In order to control for the effects of color on performance, colored images were converted to grayscale

intensity images by removing the hue and saturation information while keeping the luminance information. RGB values were converted to grayscale values by computing a weighted sum of the channels using the intensity transformation

$$I = 0.299R + 0.587G + 0.114B, \quad (3.1)$$

where  $I$  represents the grayscale intensity, and  $R$ ,  $G$ , and  $B$  corresponds to red, blue, and green channels, respectively.<sup>2</sup> Also, to control the variability in luminance and contrast levels across images, the average luminance of each image was set to 40 cd/m<sup>2</sup> and its contrast level (root-mean-square, RMS) was adjusted to 0.4. Then, the clutter was computed for each image (see the Measuring Clutter section). The distribution of clutter was similar across search images containing different target object categories (see Figure 3.2, left panel).

### 3.2.3.1.3 Labelling Stage

Images were annotated by labeling the type of potential target objects present in them. Then, target locations were marked by drawing circumscribing polygons around the target objects. The vertices of these polygons were recorded. At the end of this stage, each image was associated with an annotation consisting of: a list of target objects within the image, a list of vertices describing the circumscribing polygon for each target object, and the clutter value for the image.

### 3.2.3.1.4 Selection Stage

For each of five target categories, 800 test images were selected. The target object was present only in half of these images. Test images were chosen based on the following criteria.

---

<sup>2</sup>The weights used in conversion of RGB values to grayscale values were based on ITU-R Recommendation BT.601-7 standard for color video encoding.

First, to label images as high and low clutter, we computed the median of the clutter distribution. Images with clutter values higher than the median were marked as high clutter while images with clutter values lower than the median were marked as low clutter.

Next, we expected that target size might impact search performance in target present images. To control for any size effects, we first measured the size of each target object by computing the area of its circumscribing polygon. Target size varied depending on the target category (see Figure 3.2, right panel). For example, on average, cellphones were larger than keys. To limit the effects of unusually-sized objects, we restricted the variability in target size by including images only if  $t \in [\frac{1}{4}m, 4m]$ , where  $t$  is the target size and  $m$  is the median target size.

A final inclusion criterion considered the variants of targets. If we suspected that observers might not be familiar with a particular variant of target object, images displaying that variant were not selected. For example, in the case of cellphones, images did not include any flip-phones. Similarly, in the case of iPods, only images with iPods with a particular shape, a rectangular screen at the top and a circular area at the bottom, were included. Further, images with objects that looked highly similar to targets were also excluded. For example, images containing an iPod touch (which might look like an iPhone to the observer) were excluded. Similarly, in the case of pens, we excluded images that included makeup pencils.

At the end of this process, 800 images were selected for each target category. 400 test images were selected for the target present trials by prioritizing the amount of clutter and checking for the criteria listed above, and another 400 images without the target object were selected for the target absent trials.

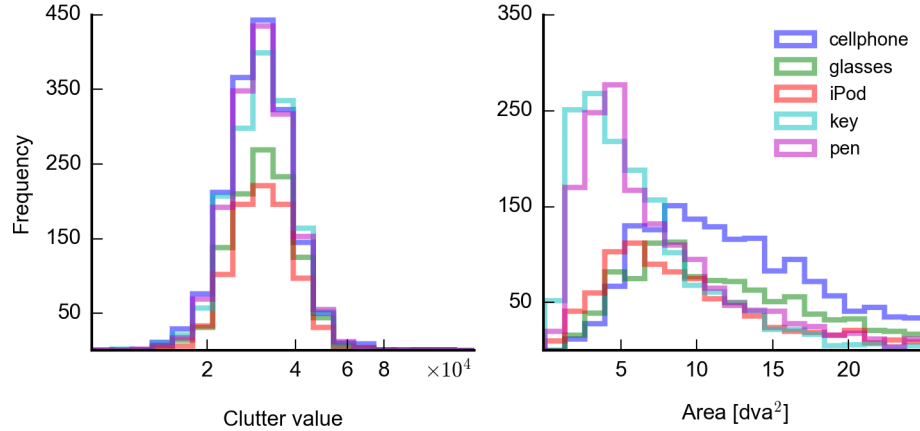


Figure 3.2: The distribution of clutter values (left panel) and target size (right panel) for each target category.

### 3.2.3.2 Preparing the Search Stimuli

For their presentation in the search task, individual images in each of the two clutter conditions were randomly assigned to either the low (5 locations) or high (13 locations) set size conditions. Potential target locations were marked by small circular cues overlaid on the image. To minimize uncertainty about cue locations, we used large cues ( $0.25^\circ$  in diameter) that were red on a grayscale image, placed the cues on a regular hexagonal grid, with spacings of  $8.0^\circ$  and  $5.9^\circ$  for set sizes 5 and 13, respectively, and made them continuously visible on the screen across each block. The average distance of the cue locations from the origin was set to be two thirds of the radius of the stimulus circle. Images were shifted and rotated so that only one of these cues appeared within the circumscribing polygon associated with the correct target location. Images were presented in a circular region,  $24^\circ$  in diameter. This region was chosen with the constraint that it contained the target object. Finally, the area around the circular region was set to uniform gray. The final form of images used as stimuli in the search task is shown in Figure 3.3.

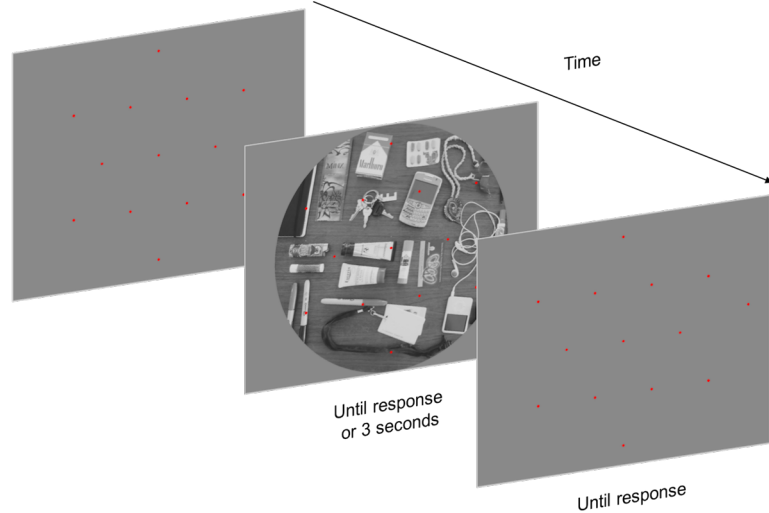


Figure 3.3: Search task sequence for a trial with keys as the search target. The small red cue markers, which were continuously visible, represent the potential target locations ( $N = 13$ ). The keys are located within the top left quadrant of the image.

### 3.2.3.3 Measuring Clutter

We quantified image clutter using a modified version of the clutter measure described in Bravo and Farid (2008). We chose this clutter measure because it has been shown to successfully predict search times in a similar set of images. Additionally, this measure is computationally efficient and scale invariant. Briefly, this measure estimates the amount of clutter in an image as a function of the relationship between the number of “segments” in an image and the scale of segmentation. The details of the segmentation procedure and our implementation of the clutter measure are described below.

#### 3.2.3.3.1 Segmentation Algorithm

To count the number segments in each image, we used the graph-based segmentation algorithm introduced by Felzenszwalb and Huttenlocher (2004). This algorithm segments the image by considering the variability of nearby regions. In

particular, it draws boundaries between regions based on pairwise comparisons of the intensities within and across regions. The threshold for drawing these boundaries is controlled by a scale parameter  $k$ . Larger  $k$  leads the algorithm to favor larger regions and results in smaller number of segments. The algorithm produces perceptually reasonable segments (e.g., see Figure 3.4) and it runs at a high speed in practice.

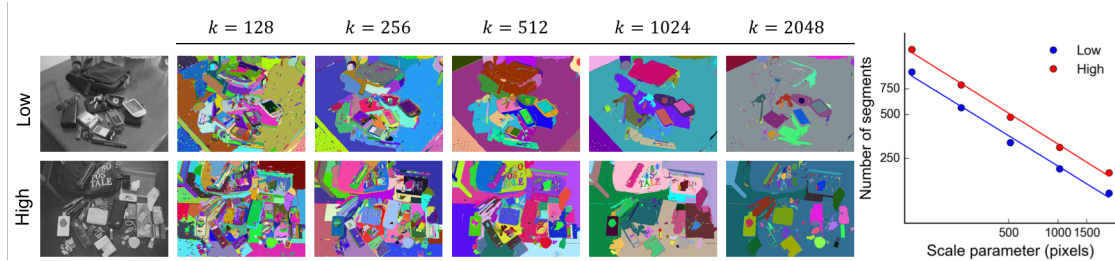


Figure 3.4: Example segmented images of a low clutter (top) and high clutter (bottom) image at 5 of the 12 possible values of the scale parameter ( $k \in [90, 4096]$ ). Color is used only to show segmented regions in the image. Plot on the right shows the number of segments as a function of the scale parameter for each image. Points represent the raw number of segments while the lines represent the log-linear fits.

Our search stimuli consisted of different cropped sections of the images for different target categories. We wanted the clutter estimates to be robust to minor changes of the position in the image. To get more stable estimates of clutter, we created random sections from the images, counted segments for each section at multiple scales, and then computed the geometric mean of segment counts across sections at each scale. At the end of this process, each image was associated with a segment count for each scale.

### 3.2.3.3.2 Clutter Measure

We measured the clutter in each image by characterizing the relationship between the scale of segmentation  $k$  and the number of segments for that scale  $y(k)$ . We determined this relationship empirically by varying the scale parameter across a

range of values, applying the segmentation algorithm, and counting the resulting number of segments. Figure 3.4 shows examples of segmented images and the number of segments at several scales of segmentation. For any given image, the number of segments is log-linearly related to the scale of segmentation, such that

$$\ln y(k) = \alpha + \beta \ln k, \quad (3.2)$$

where  $\ln$  represents the natural logarithm.

The slope of this relationship is approximately constant ( $\beta \approx -0.69$ ), but the intercept  $\alpha$  varies across images. In particular, for any setting of the scale parameter, highly cluttered images tend to have more segments than the minimally cluttered images. Therefore, we used the intercept of each image to quantify its clutter.

To get robust estimates of these log-linear relationships for each image, we (1) randomly sampled 10  $1024 \times 1024$  sections of the image, (2) computed the segment counts for each of these sections across a range of scales ( $k \in \{90, 128, 181, 256, 362, 512, 724, 1024, 1448, 2048, 2896, 4096\}$ ), and (3) computed the intercept of the log-linear fit using a least-squares procedure. The slope was computed as the average least-squares slope for all of the images in the data set ( $N = 4,953$ ), and the intercepts for individual images were fitted with this average slope held constant.

Our implementation of this clutter metric differed from the clutter metric on which it is based (Bravo & Farid, 2008) in two ways: First, we evaluated the least-squares fit in log units, where the power law functions are linear. This was done to make the model residuals more homoscedastic and thereby make the fitting more robust to outliers. Second, we sampled multiple sections from each image before segmenting them and computed the fit using the set of segment counts obtained for all of these sections. This resulted in fits that were robust to the small changes

in cropping boundaries that occurred when the images were repositioned to align the target location with the grid of cue positions.

In order to evaluate the generalizability/robustness of our clutter measurements, we also quantified clutter using alternative clutter measures including edge density (Mack & Oliva, 2004), feature congestion (Rosenholtz et al., 2005, 2007), and subband entropy (Rosenholtz et al., 2007) for the images used in our experiment. The clutter measures were all significantly correlated (see Table 3.1), suggesting that the particular choice of clutter metric is not important.

The code for implementation of the segmentation algorithm is made publicly available by its authors. A MATLAB implementation of the clutter measure using this algorithm as described above can be downloaded from our lab Github page ([https://github.com/mmmlab/clutter\\_metric\\_code](https://github.com/mmmlab/clutter_metric_code)).

Table 3.1: Pearson correlation coefficients among clutter measures.

	Segmentation	Edge density	Feature congestion	Subband entropy
Segmentation	-			
Edge density	0.67	-		
Feature congestion	0.64	0.65	-	
Subband entropy	0.57	0.70	0.60	-

Note: All  $p < 0.001$ .

### 3.2.4 Procedure

The design of the experiment was  $5$  (target object category: cellphones, glasses, iPods, keys, or pens/pencils)  $\times 2$  (relevant set size: 5 or 13)  $\times 2$  (clutter level: low or high)  $\times 2$  (target presence: target present or target absent), with one between-subjects variable (target object category) and three within-subjects variables (relevant set size, clutter level, and target presence). At the start of the search experiment, observers were randomly assigned to one of five search target categories (cellphones, glasses, iPods, keys, or pens/pencils). Observers

were instructed to detect and locate the target object within an image as quickly and accurately as possible. Additionally, they were told that if the search target was present in an image, there was only one single item or a group of items in close proximity from the search category, and the item was visible.

Before the start of each trial, observers fixated a point at the center of the display while a set of circular cues indicated the potential target locations (see Figure 3.3). Observers began the trial by pressing a start key. After the trial was initiated, the search display appeared and observers freely searched for the target. Observers were allowed three seconds to search. After either three seconds had elapsed or the observer pressed a key to complete the search early, the search image disappeared while the set of circular cues indicating the potential target locations remained on the screen. An additional cue appeared at a random location  $1^\circ$  outside the search region. Observers were instructed to make a localization decision. They fixated either the cue corresponding to the perceived location of the target (if target was present) or the additional cue (if target was absent). The cue corresponding to the current fixation was highlighted in real time to ensure that observers knew which locations they were selecting. Observers could correct their gaze if the wrong location was highlighted. When they were satisfied with their selections, observers logged their responses with a keypress.

The amount of time spent inspecting each image was recorded as the search time, and was the primary measure of performance. In target present trials, a response was registered as “correct” only if the selected cue corresponded to the target location. In the target absent trials, a response was registered as “correct” only if the absent cue location was selected. All other responses were registered as errors. Observers received auditory feedback indicating the accuracy of their responses.

Trials were blocked by the relevant set size. Each block consisted of 50 experimental trials. At the start of each block, observers completed a 13-point calibration routine covering the central  $22^\circ$  of gaze angle. The calibration was repeated until the average test-retest calibration error across gaze points fell below  $0.25^\circ$ . The calibration routine could be repeated if necessary during a block. If a blink was detected during the search phase of the trial (when the image was present on the screen), the trial was aborted, and the observer was notified. Data from aborted trials were discarded, but the image from the discarded trial was repeated later in the experiment. Observers very rarely broke fixation so that fewer than 1% of trials were aborted.

Observers completed the study in two one-hour sessions on separate days. Each session contained 8 blocks, resulting in a total of 800 trials. The block order was randomized across sessions and observers.

Observers were trained and refamiliarized with the task by completing 8 practice trials at the start of the experiment and a single practice trial at the start of each block. Data from the practice trials were excluded from the analysis.

### 3.3 Results

#### 3.3.1 Search Times

Figure 3.5 shows average search times in the target present and target absent trials. Each faint line represents data from five observers searching for one type of target (shapes) in either high clutter (red lines) or low clutter (blue lines) condition as a function of relevant set size. Each heavy line represents average search times across target categories. Two main trends are evident: (a) search times tend to increase as the relevant set size increases, and (b) search times tend to increase as the amount of clutter increases.

Search times were analyzed by conducting a  $5$  (target object category: cell-phones, glasses, iPods, keys, or pens/pencils)  $\times 2$  (relevant set size: 5 or 13)  $\times 2$  (clutter level: low or high)  $\times 2$  (target presence: target present or target absent) mixed design ANOVA, with one between-subjects variable (target object category) and three within-subjects variables (relevant set size, clutter level, and target presence).

The ANOVA revealed main effects of clutter level,  $F(1, 20) = 260.37, p < 0.001$ , of relevant set size,  $F(1, 20) = 48.69, p < 0.001$ , and of target presence,  $F(1, 20) = 173.69, p < 0.001$ . In particular, average search times were longer in the high clutter condition ( $M = 1.37, SE = 0.01$ ) than in the low clutter condition ( $M = 1.18, SE = 0.01$ ), suggesting that clutter degrades search performance. Search times were also longer in the set size 13 condition ( $M = 1.37, SE = 0.01$ ) than in the set size 5 condition ( $M = 1.19, SE = 0.01$ ), confirming our manipulation of set size. Finally, target absent trials resulted in longer search times ( $M = 1.55, SE = 0.01$ ) than the target present trials ( $M = 1.00, SE = 0.01$ ). The main effect of target category did not reach significance,  $F < 1, n.s.$

Our analysis also revealed several significant interaction effects. There was a significant clutter level  $\times$  relevant set size interaction,  $F(1, 20) = 4.93, p = 0.036$ , which suggests that the effect of clutter tends to be larger for larger set size. The clutter level  $\times$  target category interaction was also significant,  $F(4, 20) = 5.29, p = 0.005$ , suggesting that the effect of clutter was larger for some target categories than others. Additionally, clutter level  $\times$  target presence interaction reached significance,  $F(1, 20) = 48.65, p < 0.001$ , suggesting larger effects of clutter in the target absent trials compared to the target present trials. Moreover, set size  $\times$  target presence interaction was significant,  $F(1, 20) = 11.01, p = 0.003$ , which suggests larger set size effects in the target absent trials compared to the target present trials.

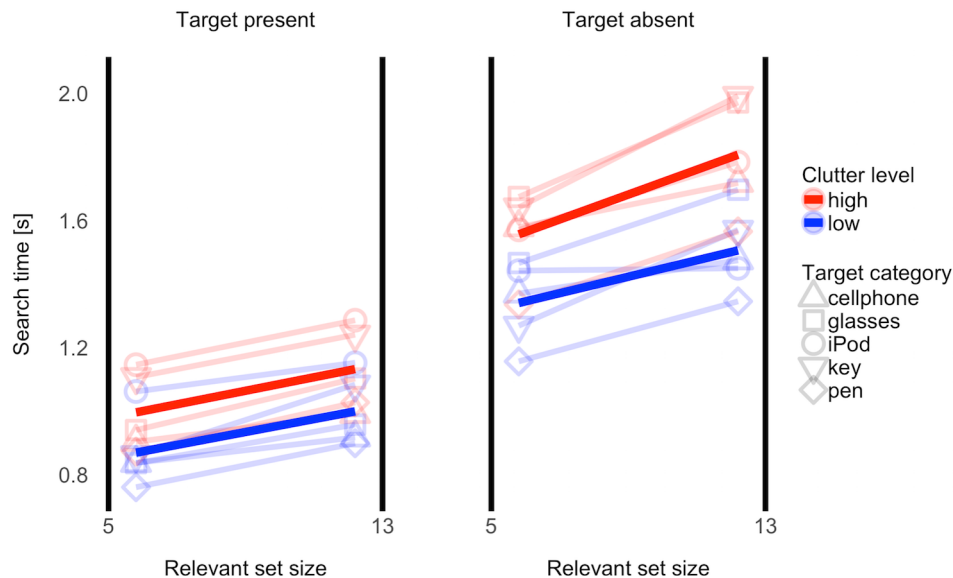


Figure 3.5: Average search times as a function of relevant set size in the target present trials (left panel) and in the target absent trials (right panel). Each combination of line and symbols represents data from five observers searching for one type of target (shapes) in either high clutter (red lines) or low clutter (blue lines) condition. Average search times across target categories are represented by the heavy lines.

### 3.3.2 Fixation Distributions

As a further check on our manipulation of set size, we examined observers' fixation distributions during search. If observers make use of the target location information provided by the cues when planning their fixations, they should be more likely to fixate the cued locations than other locations in the display. Figure 3.6 shows the fixation distributions for each of the set size conditions, aggregated across all observers and trials, with the first and last fixations excluded. We excluded the first fixation from the analysis because observers always started the search by fixating at the center of the screen, and we excluded the last fixation from the analysis to avoid biasing the fixation distributions towards the target locations (i.e., since observers typically completed the search by fixating the target location). Our analysis of fixation distributions shows that observers indeed use

cue location information when selecting their fixation locations, confirming the effectiveness of our set size manipulation.

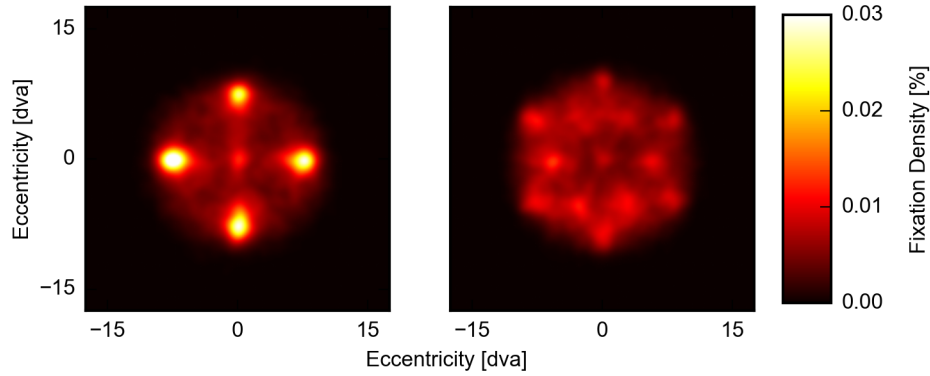


Figure 3.6: Aggregated fixation distributions across all of the observers, for set size 5 (left panel) and for set size 13 (right panel). The first and final fixations were excluded from the analysis.

### 3.3.3 Search Target Sizes

We also examined how search time changed as a function of target size. Although we restricted the size of the targets to a limited range, there was still some degree of variability. Each target's size was quantified using either the area of its circumscribing polygon or the length of the longest axis of this polygon. To remedy the curvilinear relationship observed between the target area and the search times, the areas were transformed by taking their square root, which resulted in a more linear relationship. Figure 3.7 shows that (a) search times tend to decrease as the search target gets larger in size, and (b) some targets are larger, on average, than others. The analysis showed that search times decrease significantly as target size increases, both when the size was measured as the area ( $r = -0.29, p < 0.001$ ) and when it was measured as the length of longest axis ( $r = -0.35, p < 0.001$ ). These results suggest that target size may be one of the factors driving differences in search performance among target categories.

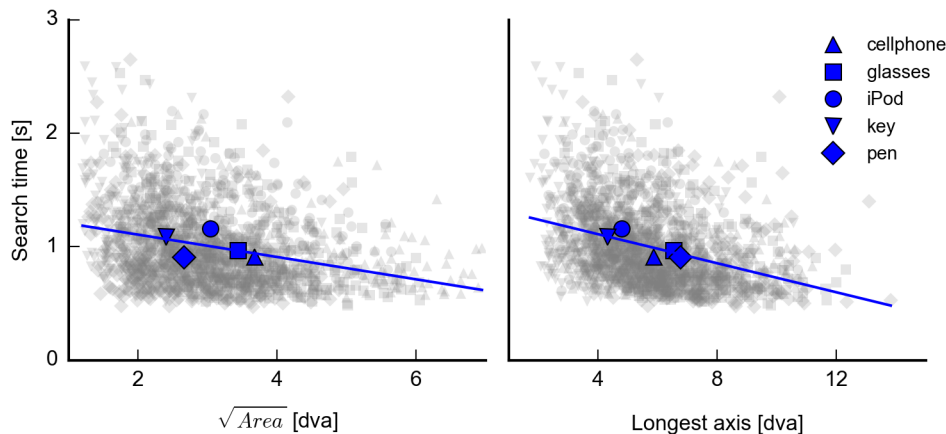


Figure 3.7: Search times as a function of target size represented by the square root of the area (left panel) or the longest axis (right panel) of its bounding polygon. Each gray dot represents the average search time across five observers for a particular target in an image. Shaped markers represent the average size for each target category. Blue lines represent the least-squares linear fits.

### 3.3.4 Search Target Categories

Our stimulus set contained some common images across different target categories. That is, in some cases, different observers searched for different targets in the same image. These cases gave us the ability to dissociate effects of the search image from those of the search target and to directly examine the effect of target category on search performance. Figure 3.8 shows the average search times while searching for different targets in the same image. For example, the first plot shows the search time while looking for a cellphone compared to the search time while looking for the other targets in the same image. If the search performance was only determined by the amount of clutter or the relevant set size, then all points would line up on the diagonal. However, these results show that some targets were more difficult to find than others. For example, on average, observers seem to be faster at locating cellphones than other targets. We discuss potential implications of these results in the Discussion section.

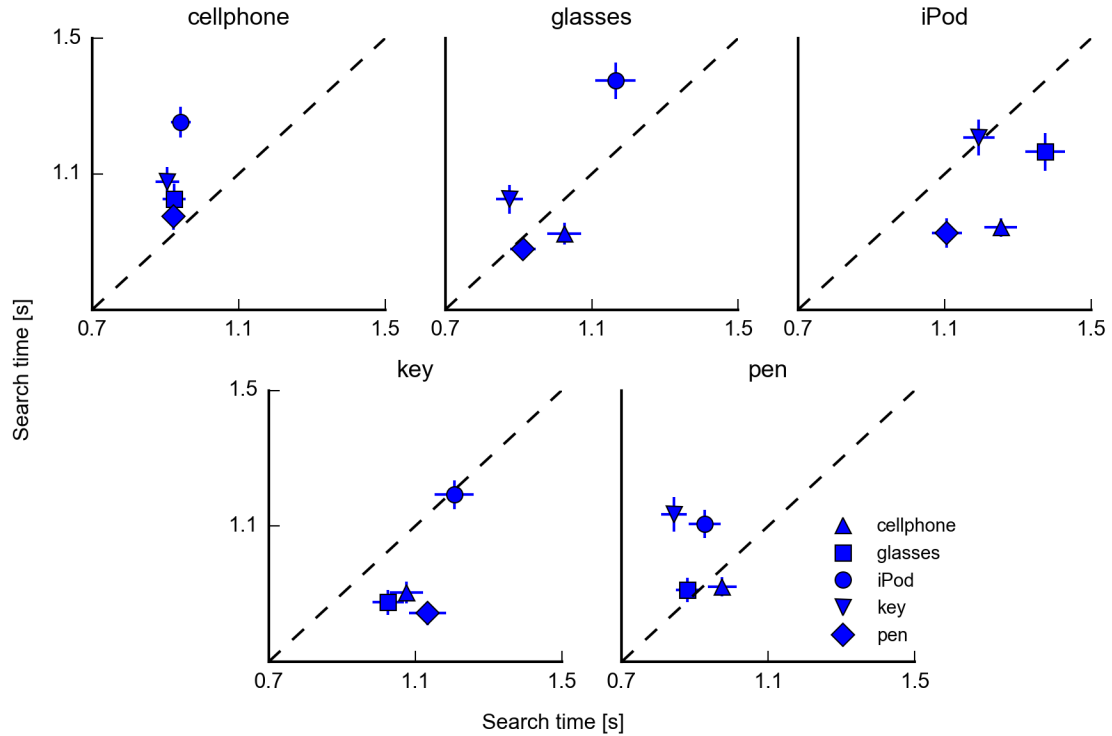


Figure 3.8: Average search times for different targets in common images. Each panel compares search time for a particular target category (on the x-axis) to search time for other targets (on the y-axis) in the same image. Each point represents average search times across all images that contained the indicated pair of targets. Error bars indicate standard error.

### 3.3.5 Error Rates

Error rates were defined as the proportion of trials where observers did not fixate the correct cue location (see the Procedure section). The probability of choosing any location other than the correct location (i.e., the baseline error rate) was computed as 0.83 and 0.93 for set sizes of 5 and 13, respectively. Table 3.2 shows error rates across conditions. In general, error rates were well below baseline rates, suggesting that observers were extremely accurate in their judgments.

Table 3.2: Error rates across conditions.

Clutter level	Relevant set size	Target category					average
		cellphone	glasses	iPod	key	pen	
Low	5	0.04	0.03	0.09	0.03	0.06	0.05
	13	0.05	0.06	0.11	0.04	0.09	0.07
High	5	0.03	0.05	0.09	0.06	0.09	0.06
	13	0.05	0.07	0.12	0.09	0.11	0.09

### 3.4 Discussion

The purpose of the current study was to determine how clutter affects search for categorical targets in real-world scenes. In particular, we sought to disentangle the effects of extrinsic position uncertainty (i.e., search set size) from those due, through the modulating effect of clutter, to intrinsic position uncertainty (Semizer & Michel, 2017). Our results exhibited several trends:

First, search times increased significantly as a function of increasing clutter. This pattern was evident across target categories, but the effect was larger for some targets than others. Second, search times increased significantly as the number of possible target locations increased, revealing the classic set size effect. This finding provided evidence that our manipulation of extrinsic uncertainty, via the relevant set size, was successful. Third, when all other manipulated factors (i.e., clutter level and set size) were fixed, search times changed as a function of target category. Finally, search times decreased significantly as a function of target size, both when the size was measured as the area of the circumscribing polygon and when it was measured as the length of the longest axis of this polygon. We discuss potential implications of these findings below.

### 3.4.1 Clutter-Specific Effects

Several studies have shown that clutter degrades search performance in naturalistic stimuli (e.g., Bravo & Farid, 2004, 2008; Henderson et al., 2009; Neider & Zelinsky, 2011; Rosenholtz et al., 2007). However, there are various ways in which clutter can lead to the observed performance impairments. For example, clutter has been used as a proxy for set size in natural scenes because, as scene clutter increases, the (implicit) number of potential target locations also tends to increase (Rosenholtz et al., 2005, 2007). Additionally, clutter can force observers to consider features at irrelevant locations during search, exacerbating the effects of intrinsic position uncertainty (Michel & Geisler, 2011; Semizer & Michel, 2017). As a result, localizing the source of peripherally perceived stimuli becomes more difficult, which degrades search performance. Finally, clutter can make search more difficult by obscuring search targets. Adding clutter to real-world scenes increases the probability that objects will partially or completely occlude one another. In the current study, we controlled for set size and for occlusions of the search target to isolate those effects of clutter that are due to intrinsic position uncertainty.

The results of the current study, obtained using real-world images, are in broad agreement with those of a related study that showed how clutter degrades search performance in synthetic noise displays (Semizer & Michel, 2017). However, the results of the current study differ in one notable respect. Semizer and Michel (2017) reported that the effect of extrinsic position uncertainty diminished at larger set sizes when the searcher was limited by intrinsic position uncertainty. As a result, search performance was similar across cluttered and uncluttered conditions when the relevant set size was large. However, our results showed that search performance was worse in the high clutter condition than in the low clutter

condition regardless of the relevant set size. This difference might be due to either of two reasons: First, the images in our experiment were far less cluttered than the synthetic displays created in the lab. When measured using the same clutter metric, the synthetic stimuli from Semizer and Michel (2017) yielded clutter values of around  $\alpha = 4 \times 10^8$ , which was several orders of magnitude larger than the clutter values measured for our images (see Figure 3.2). Second, the relevant set sizes used in our study were much smaller than those in Semizer and Michel (2017). In the current study, the set sizes consisted of either 5 or 13, while the set sizes of Semizer and Michel (2017) ranged from a minimum of 37 to a maximum of 817 potential target locations. Indeed, our results are completely consistent with those of Semizer and Michel (2017) when we consider only the smaller set sizes used in that study.

### 3.4.2 Characterizing Set Size

In many traditional search experiments, set size is defined as the number of items in a search display (see Wolfe, 1998, for a review). This type of set size is also called the “display set size”. A task-relevant subset of these items form the “relevant set size”, which can be manipulated independently of the display set size by cuing only the locations that might contain the target (Palmer, 1994, 1995). Palmer (1994) introduced this distinction to characterize searches of displays comprising a small number of elements on a uniform background. However, the notion of relevant set size is especially critical in searches for which the number of elements is either very large or undefined. For example, researchers have used this method to define and manipulate set sizes for searches in synthetic noise displays (Burgess & Ghandeharian, 1984b; Eckstein et al., 2007; Manjeshwar & Wilson, 2001; Najemnik & Geisler, 2005; Semizer & Michel, 2017; Swensson & Judy, 1981) and in structured medical images (Bochud et al., 2004; Eckstein &

Whiting, 1996). In the current study, we likewise used Palmer’s (1994) method, cueing potential target locations to define the relevant set size in natural scenes.

In the context of natural scenes, there are several ways in which the effective set size might be made functionally smaller than the (nominal) relevant set size. First, if observers only consider the locations that contain “stuff” (i.e., objects, items, or feature elements) and preferentially fixate only cues that fall on these locations, this would reduce the relevant set size. However, in the context of natural scenes, characterizing what “stuff” entails is problematic because labeling or counting every single item in a scene is an ill-defined problem. In particular, identifying what constitutes an object or a background is not clear (Neider & Zelinsky, 2008, 2011; Rosenholtz et al., 2007; Wolfe, Võ, Evans, & Greene, 2011), especially when texture elements are involved. For example, in a kitchen scene, if objects are placed on a table covered with a patterned cloth or if objects were placed on other objects, it is not clear how to segment the scene into object or background. Similarly, if a scene contains a textbook (with text or illustrations on its cover), or a patch work quilt, or an articulated figure, it is ambiguous at what level objects should be segmented (should individual letters be considered objects?, individual patches?, individual parts?).

In the current study, if observers preferentially fixated cues that fall on objects, we would expect this to reduce the effective set size similarly for both clutter conditions. Since both set sizes would be reduced, this should not systematically influence the set size and clutter effects observed in our experiment. To test this relationship empirically (and based on a reviewer’s suggestion), we characterized the “effective” set size by counting the number of object cues (cues that fall on objects) in each trial of our experiment. The proportion of object cues were similar across low and high clutter conditions, 0.72 and 0.79, respectively; suggesting no substantial differences between the clutter conditions. Therefore, if having cues

land on the “background” reduced the effective set size, then it did so similarly for both clutter conditions. Figure 3.9 shows search times as a function of the number of object cues, suggesting that the proportion of object cues is independent of clutter in this set of stimuli.

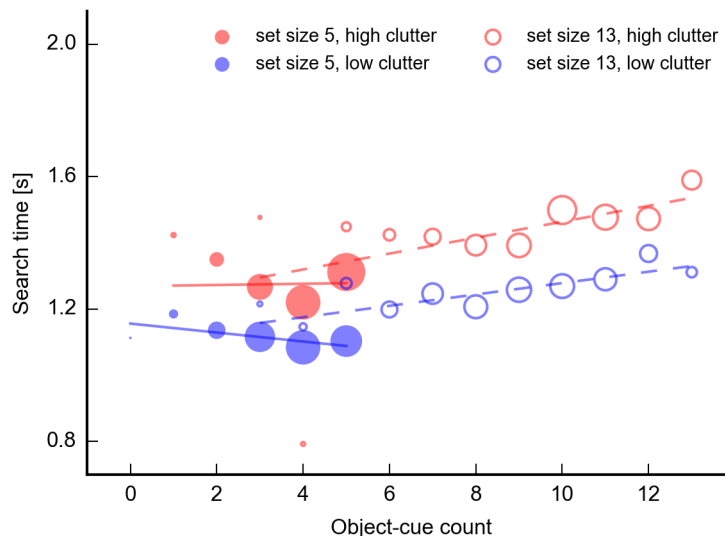


Figure 3.9: Search times as a function of the object-cue count (i.e., the number of cues that fall on objects) in the search image. The area of each marker is proportional to the number of trials exhibiting the corresponding object-cue count. Solid and dashed lines represent the least-squares linear fits for set sizes 5 and 13, respectively.

Set size might also be effectively reduced when targets are not distributed uniformly across cued locations. If the appearance of the target at a subset of the cued locations were much more probable than at other locations, then observers might restrict their searches to those locations within the probable subset, effectively reducing the set size. To investigate this possibility, we measured the likelihood that each cue location contained a target in our stimulus set. Target locations were well distributed across possible locations (Figure 3.10), with the exception of the center location, which was somewhat underrepresented.

To quantify the effect of this non-uniformity of the target location distributions in reducing the relevant set size, we computed the information entropy (Shannon,

1948) for each set size, given by

$$H = - \sum_i^n p_i \log_2 p_i, \quad (3.3)$$

where  $p$  is the probability that a cue location contains a target and  $n$  is the relevant set size. If the search targets were uniformly distributed across the cued locations, then the entropies associated with set sizes 5 and 13 would be 2.32 and 3.70 bits, respectively. The corresponding entropies computed for the empirical distributions measured in our experiment (Figure 3.10) were 2.21 and 3.59 bits, respectively. This reduction in entropy was small for both set sizes (approximately 5% for set size 5 and 3% for set size 13) suggesting that any reduction in effective set size caused by the non-uniformity of the target distributions should have a negligible effect on search performance.

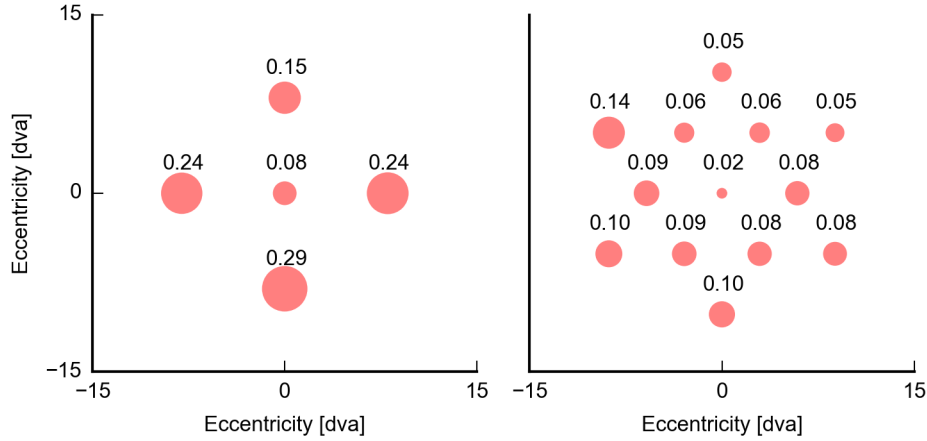


Figure 3.10: Distribution of target locations for set size 5 (left) and set size 13 (right) conditions. The area of each location marker is proportional to the empirical target probability for the corresponding location. The probabilities are also printed above each cue location marker.

### 3.4.3 Target-Specific Effects

We used different types of target categories in the search experiment and this allowed us to examine how a particular target category contributes to the effect of clutter on search performance. Our analysis of search times suggest that target category interacts with clutter. But what makes a target category more or less susceptible to clutter? Intuitively, it seems obvious that certain features of the target (e.g., target size, color, shape, etc.) might interact with features of clutter to determine search performance. Imagine, for example, a peripheral search task that requires the localization of a target object among green distractors. If the target object is red, then green objects will not provide effective clutter because their features are not confusable with features of the target. However, if the target is green, then the distractors should provide effective clutter.

More generally, the similarity of the target to features of the background might affect the susceptibility to clutter. Search times are longer when targets are similar to distractors or when distractors are dissimilar to other distractors (J. Duncan & Humphrey, 1989). Also, the similarity of targets to the search background affects search performance (Neider & Zelinsky, 2006). For example, when the search target and the background share a common spatial frequency band, search becomes more difficult (Semizer & Michel, 2017). Thus, when investigating the effects of clutter on search performance, it makes sense to expect performance differences depending on the similarity of target features to background features. For example, imagine a scene of leaves. The traditional models of clutter would consider this scene “highly cluttered” and predict poor search performance in this scene regardless of the type of the search target. However, if the search target, such as a cellphone, does not share a lot of similar features with the background, then the search should be pretty easy. In fact, there are models of clutter from

the field of image optics, which quantify clutter in electro-optical images in terms of the target-background similarity (e.g., Chang & Zhang, 2006; Chu et al., 2012; Moore, Camp, Moyer, & Halford, 2010b; Schmieder & Weathersby, 1983; Silk, 1995; Tidhar et al., 1994).

Another feature of the target that might affect its susceptibility to clutter is its size. Target size has been identified as one of the fundamental attributes in guiding attention (Wolfe & Horowitz, 2004). As one of the target-specific features, we measured size of the search targets in our stimuli and examined how search performance changes as a function of target size. The analysis suggests that larger targets are associated with shorter search times. This means that, within the context our study, larger targets are easier to find. Similarly, target size within a category might affect susceptibility to clutter, and the strength of the relationship between clutter and search performance might depend on details particular to different search targets. Revealing the nature of these specific target features remains an open question, and one that we plan to pursue in future work.

### 3.5 Conclusion

Overall, our results demonstrate that increased clutter reduces performance in searches of real-world scenes, and does so independently of set size. When considered in the context of previous studies, Michel and Geisler (2011) and Semizer and Michel (2017), that explicitly modeled intrinsic position uncertainty, this study suggests that the intrinsic position uncertainty of peripheral vision significantly limits searches of real world scenes in the same way it limits searches of synthetic scenes. Therefore, it is important to account for these effects of intrinsic position uncertainty when evaluating and modeling performance in search tasks.

## Chapter 4

### Study II

#### 4.1 Introduction

Visual search becomes more difficult when the visual environment in which we perform the search is *cluttered* (Bravo & Farid, 2004, 2008; Henderson et al., 2009; Lohrenz et al., 2009; Neider & Zelinsky, 2011; Rosenholtz et al., 2007). The effects of clutter on performance has been investigated in various tasks by several research domains, such as psychology, engineering, and computer science; therefore, there are various ways defining and measuring clutter. The most naïve way of defining clutter is related to the number of objects in a display. This definition of clutter is similar to the notion of set-size (Palmer, 1994, 1995). In traditional visual search studies, as the number of distractors in a display increases, search performance decreases (Wolfe, 1994). Similarly, the density of items in naturalistic scenes correlates with the amount of clutter in the display (van den Berg et al., 2009). Although this definition of clutter is intuitive, it may not be easy, or even possible, to estimate the number or density of items in natural scenes, where identifying clearly defined objects is difficult (Neider & Zelinsky, 2008, 2011; Rosenholtz et al., 2007; Wolfe et al., 2011).

Clutter can also be defined in terms of the layout of items in a display. According to this account, clutter is related to the arrangement, structure, or density of items in a scene (Bravo & Farid, 2008; Rosenholtz et al., 2005; van den Berg et al., 2009). For example, clutter is associated with high variability of luminance,

contrast, orientation, and color in a display (Rosenholtz et al., 2005), large number of regions in a display (Bravo & Farid, 2008), or high density of edges in a display (Mack & Oliva, 2004).

Another way of defining clutter is related to the target-background (or target-distractor) similarity in a display. Performance in visual search tasks using structured stimuli decreases as target-background similarity increases (J. Duncan & Humphrey, 1989; Wolfe et al., 2002). Models from the field of image optics define clutter in terms of the similarity between target and background features (Camp et al., 2010; Chu et al., 2012; He et al., 2008).

These accounts of clutter view clutter primarily as a property of the display. Clutter is also defined in terms of task performance. According to this perspective, clutter refers to “the state in which excess items, or their representation or organization, lead to a degradation of performance at some task” (Rosenholtz et al., 2005, p. 761). The structure or density of items in a display is also accounted for by this view of clutter.

The particular choice of clutter definition motivates the approach in measuring clutter in a display. Approaches to measuring clutter include developing objective clutter metrics, manipulating the amount of clutter in a display and measuring performance, and collecting subjective judgments based perceived clutter. The most common objective clutter measures in psychophysics include edge density (Mack & Oliva, 2004), the scale invariant clutter measure (Bravo & Farid, 2008), feature congestion (Rosenholtz et al., 2005, 2007), subband entropy (Rosenholtz et al., 2007), the proto-object model (Yu et al., 2014), and the color-clustering clutter (C3) model (Lohrenz et al., 2009). These measures suggest that as the amount clutter in a scene increases, search performance decreases (Bravo & Farid, 2008; Henderson et al., 2009; Mack & Oliva, 2004; Rosenholtz et al., 2007; Neider

& Zelinsky, 2011). Although these measures successfully predict search performance, they do not account for the properties of search targets. In particular, they do not consider target-background similarity in quantifying the effect of clutter on search performance. However, when search targets and background share similar features (e.g., spatial frequency content), search performance declines (Semizer & Michel, 2017).

There are other clutter measures, in the field of image optics, which quantify clutter using the target-background similarity. These measures include the statistical variance (SV) metrics (Schmieder & Weathersby, 1983; Silk, 1995), the probability of edge detection (POE) metric (Tidhar et al., 1994), the target structure similarity (TSSIM) metric (Chang & Zhang, 2006), the contrast sensitivity function (CSF) based metric (Chu et al., 2012), and the masked target transform volume (MTTV) metric (Moore et al., 2010a, 2010c). Most of these metrics account for target-background similarity but only at the exemplar level. They can not predict search performance for categorical targets or for target absent cases.

In this study, we propose two new clutter metrics based on two types of target-background similarity (i.e., exemplar level and category level) to predict the effect of clutter on search performance. The exemplar clutter metric quantifies the overlap between features of a search target that is present in an image (at the exemplar level) and features of a search background. The category clutter metric quantifies the overlap between features of a search target category (at the categorical level) and features of a search background. The category clutter metric is used to predict search performance in target absent images while both the category clutter metric and the exemplar clutter metric are used to predict search performance in target present images. We test the predictive performance of these metrics using a set of search data and compare their performance to an existing target-agnostic clutter metric.

## 4.2 Methods

To quantify the effect of target-background similarity on search performance, we propose two new metrics. In a nutshell, our metrics measure similarity with respect to features (i.e., the orientation subbands) of an image while also accounting for the size of a search target. The first metric, the exemplar clutter metric, operates at the exemplar level. This metric quantifies similarity as a function of features shared between a search target that is present in an image and the search background. Since this metric requires presence of a target object in an image, it can be used to predict search performance only in target present images.

While searching for categorical targets, we would expect to see an effect of target category even if the particular search image does not contain an instance of the target category. Thus, we propose a second metric, the category clutter metric, to measure target-background similarity at the categorical level. This metric quantifies similarity as a function of features shared between a search target category and the search background. Since this metric does not require presence of a target object in an image, it can be used to predict search performance in target absent images as well as in target present images.

To test whether representing the target-background similarity meaningfully contributes to predicting search performance, we compared performance of these two metrics to an existing metric which is agnostic to the target-background similarity. In particular, we used the clutter measure developed by Bravo and Farid (2008) as the target-agnostic clutter metric. We chose this metric because it has been shown to predict search performance using a similar set of images, it is computationally efficient, and scale-invariant.

Below, we describe the exemplar clutter metric, the category clutter metric, and the target-agnostic clutter metric in more detail.

### 4.2.1 The Exemplar Clutter Metric

The exemplar clutter metric measures the overlap between search target and search background features at the exemplar level. It requires an instance of search target object to be present in an image. Below, we describe our approach in representing a search target and search background followed by the method of quantifying similarity.

#### 4.2.1.1 Representing the Search Target

We represented a search target object in an image in the following steps: First, we created a patch with the target object centered at the patch. The size of the patch was set to the longest dimension of the bounding box enclosing the target object. Then, we constructed a steerable pyramid representation of the target patch using 3 different scales and 4 different orientations, and extracted the orientation subbands. A steerable pyramid is an image decomposition method which creates a linear decomposition of an image into multiple scale and orientation subbands (Simoncelli & Freeman, 1995). Using steerable pyramids with certain number of orientations is computationally less expensive than computing Fourier transform at all orientations.

Next, to minimize any effects of the remaining background information in the target patch, we windowed the patch responses using a 2D raised cosine window (Hanning window). At the end of this process, each target patch was associated with a single normalized feature vector.

#### 4.2.1.2 Representing the Search Background

Similar to the search target, we processed the search background in the following steps: First, we constructed a steerable pyramid representation of the entire

image using 3 different scales and 4 different orientations. Then, using patches with the same size as the target object, we computed subband responses at the patch centered at each pixel in the image. At the end of this process, each patch in the image was associated with a normalized feature vector.

#### 4.2.1.3 Measuring Similarity

After computing feature vectors associated with the search target and the search background, we measured the similarity between these features. We used the Mahalanobis distance (Mahalanobis, 1936) as a measure of similarity, which can be expressed as

$$D_M = \sqrt{(t - \mu)^T \Sigma^{-1} (t - \mu)} \quad (4.1)$$

where  $t$  is the normalized target feature vector, and  $\mu$  and  $\Sigma$  are the mean and covariance of the normalized background feature vectors. Larger values of  $D_M$  indicate lower similarity. Finally, the exemplar clutter metric is defined as

$$C_E = \frac{1}{\sqrt{D_M}}, \quad (4.2)$$

where larger values of  $C_E$  indicate higher clutter.

### 4.2.2 The Category Clutter Metric

The category similarity metric measures the overlap between search target and background features at the categorical level. It does not require an instance of search target object to be present in an image; thus, it can be used to predict search performance for categorical targets in target absent images. Below, we describe our approach in representing a categorical search target and search background followed by the method of quantifying similarity.

#### 4.2.2.1 Representing the Categorical Search Target

In the exemplar similarity metric, search targets were represented using features of targets present in images. Forming a categorical representation of search targets is not trivial, especially since we do not have a single target object representative of a category. One possible solution involves taking the average across features of instances present in images. However, this method would result in a loss of information which would probably lead to a poor representation of a target category.

Therefore, we represented a categorical search target using a distribution of feature vectors. We constructed these distributions using the search target objects present in the images. Since sizes of each search target object varied, we had to come up with a measure to represent the typical size of targets belonging to a category. We computed the median size of the targets per category. The median size served as the patch size to which each target patch was resized. Then, we extracted the feature vectors by building pyramids, and approximated the joint distribution of these feature vectors as a Gaussian distribution.

#### 4.2.2.2 Representing the Search Background

We represented the search background in a similar way as in the exemplar clutter metric; however, this time we used the median size of targets per category as the patch size while windowing the background. We processed the background image as described in Section 4.2.1.2. Then, we approximated the joint distribution of feature vectors as a Gaussian distribution.

#### 4.2.2.3 Measuring Similarity

We compared the categorical search target distribution and the background distribution by measuring the Kullback-Leibler (KL) divergence (Kullback & Leibler, 1951) between two multivariate Gaussian distributions, given by

$$D_{\text{KL}} = \int_{-\infty}^{\infty} \log \left( \frac{p}{q} \right) p, \quad (4.3)$$

where  $p$  and  $q$  are categorical search target distribution and the background distribution of feature vectors, respectively. Larger values of  $D_{\text{KL}}$  indicate lower similarity. To avoid overestimation while measuring similarity in a given image, feature vectors corresponding to the particular target object present in the current image are left out of the target distribution. Finally, the category clutter metric is defined as

$$C_{\text{C}} = \frac{1}{\sqrt{D_{\text{KL}}}}, \quad (4.4)$$

where larger values of  $C_{\text{C}}$  indicate higher clutter.

#### 4.2.3 The Target-Agnostic Metric

To assess the performance of our new metrics, we used a target agnostic metric; i.e., a version of the clutter metric developed by Bravo and Farid (2008). In brief, this metric is based on an algorithm (Felzenszwalb & Huttenlocher, 2004) which segments images at multiple scales. Clutter is measured using a least-squares procedure describing the relationship between the number of segments and the scale of segmentation. Larger values of this metric indicate higher clutter. Details of the particular implementation of this metric can be seen in Chapter 3.

### 4.3 Results

We computed the exemplar clutter metric, the category clutter metric, and the target-agnostic metric for images used in Study I. To investigate the performance of the clutter metrics in predicting search performance, we used the search time data from Study I, and compared the performance of clutter metrics to the performance of the target-agnostic clutter metric. Details of the search task and images can be seen in Chapter 3. Some of these images had targets that were too small to build a steerable pyramid with 3 levels. These images corresponded to less than 1% trials and we excluded these trials from the analysis.

To examine the performance of these metrics in predicting search performance, we fitted Generalized Linear Models (GLMs) to the data using a  $\text{Gamma}(\alpha, \beta)$  distribution, suggested by the distribution of search times. To evaluate the goodness-of-fit, we calculated the Akaike information criterion, AIC (Akaike, 1973), as well as a pseudo  $R^2$  (McFadden, 1974), given by

$$R^2 = 1 - \frac{\ln L(M_{Full})}{\ln L(M_{Intercept})}, \quad (4.5)$$

where  $L$  is the likelihood,  $M_{Full}$  is the model with predictors, and  $M_{Intercept}$  is the model without predictors.

#### 4.3.1 The Exemplar Clutter Metric

The exemplar clutter metric was computed only for target present images because it requires a target object to be present in an image. Figure 4.1 shows search times as a function of the exemplar clutter metric and the target agnostic metric. GLM results showed that while both metrics significantly predicted search performance, the exemplar clutter metric outperformed the target-agnostic metric,

indicated by smaller AIC and larger  $R^2$  (see Table 4.1).

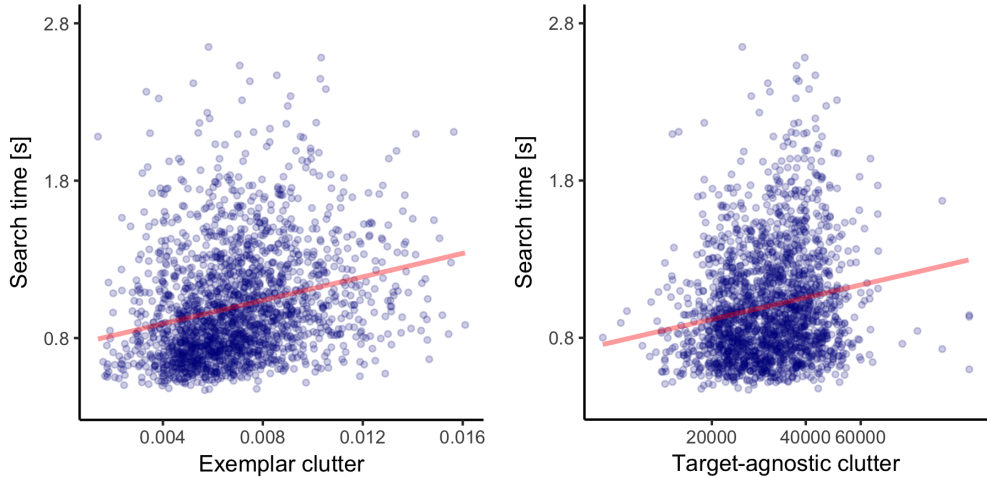


Figure 4.1: Search times as a function of the exemplar clutter metric (on the left) and the target-agnostic metric (on the right) in target present images. Lines represent GLM fits.

Table 4.1: Parameter estimates, AICs, and pseudo  $R^2$ s for target present images.

Metric	$\alpha$	$\beta_0$	$\beta_1$	AIC	$R^2$
Exemplar clutter	9.77	0.74	37.28	959.51	0.13
Category clutter	9.64	0.81	0.74	985.30	0.11
Target-agnostic clutter	9.42	-1.05	0.12	1033.06	0.06

Note: All  $p < 0.001$ .  $\alpha$  is the shape parameter and  $\beta$ s are the regression coefficients.

### 4.3.2 The Category Clutter Metric

The category clutter metric was computed both for target present and target absent images. Below, we discuss results for target present images followed by the results for target absent images.

#### 4.3.2.1 Target Present Images

Figure 4.2 shows search times as a function of the category clutter metric in target present images. GLM results showed that the category clutter metric significantly predicted search performance. When compared with the target-agnostic metric, the category clutter metric resulted in a smaller AIC and larger  $R^2$ , suggesting a better performance for the category clutter metric in predicting search performance (see Table 4.1).

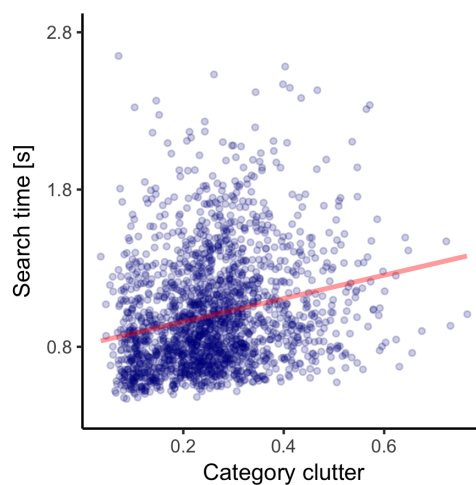


Figure 4.2: Search times as a function of the category clutter metric in target present images. Lines represent GLM fits.

#### 4.3.2.2 Target Absent Images

Figure 4.3 shows search times as a function of the category clutter metric and the target-agnostic metric for target absent images. GLM results showed that both metrics significantly predicted search performance. The target-agnostic clutter metric resulted in a smaller AIC and larger  $R^2$  than the category clutter metric, suggesting a better performance for the target agnostic metric in predicting search performance (see Table 4.2).

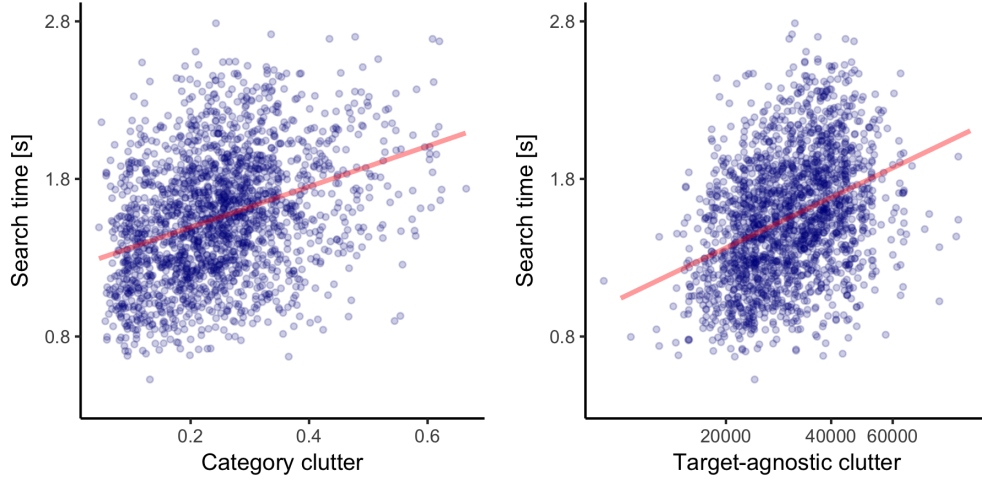


Figure 4.3: Search times as a function of the category clutter metric (on the left) and the target-agnostic metric (on the right) in target absent images. Lines represent GLM fits.

Table 4.2: Parameter estimates, AICs, and pseudo  $R^2$ s for target absent images.

Metric	$\alpha$	$\beta_0$	$\beta_1$	AIC	$R^2$
Category clutter	16.77	1.24	1.29	1709.78	0.12
Target-agnostic clutter	16.99	-3.21	0.46	1683.08	0.14

Note: All  $p < 0.001$ .  $\alpha$  is the shape parameter and  $\beta$ s are the regression coefficients.

### 4.3.3 Contributions of Clutter Metrics

To directly investigate the contributions of clutter metrics in explaining search performance, we fitted nested models and compared the performance of these models using the Likelihood Ratio Test, separately for target present images (see Tables 4.1) and for target absent images (see Tables 4.2).

For target present images, we started with a model that includes the target-agnostic metric as the only predictor, and iteratively added the other clutter metrics. Adding the exemplar clutter metric to this model significantly improved the model fit ( $\chi^2 = 12.89, p < 0.001$ ). Then, adding the category clutter metric to the model led to further improvement of model fit ( $\chi^2 = 2.27, p < 0.001$ ).

Similarly, for target absent images, we started with a model that includes the target-agnostic metric as the only predictor. Adding the category clutter metric to the model significantly improved the model fit ( $\chi^2 = 7.15, p < 0.001$ ).

Table 4.3: Parameter estimates, AICs, and pseudo  $R^2$ s in target present images for nested models.

Predictors	$\alpha$	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	AIC	$R^2$
Target-agnostic	9.42	-1.05	0.12			1033.06	0.06
Target-agnostic + Exemplar	10.02	2.97	-0.168	-32.92		909.86	0.18
Target-agnostic + Exemplar + Category	10.13	2.62	-0.13	-27.78	-0.33	889.04	0.20

Note: All  $p < 0.001$ .  $\alpha$  is the shape parameter and  $\beta$ s are the regression coefficients.

Table 4.4: Parameter estimates, AICs, and pseudo  $R^2$ s in target absent images for nested models.

Predictors	$\alpha$	$\beta_0$	$\beta_1$	$\beta_2$	AIC	$R^2$
Target-agnostic	16.99	-3.21	0.46		1683.08	0.14
Target-agnostic + Category	18.07	-2.30	0.35	0.95	1559.85	0.20

Note: All  $p < 0.001$ .  $\alpha$  is the shape parameter and  $\beta$ s are the regression coefficients.

## 4.4 Discussion

The purpose of the current study was to develop new clutter metrics which account for the target-background similarity in predicting search performance. The target-background similarity was defined both at the exemplar and at the categorical level. The exemplar clutter metric measured similarity by quantifying shared features between a search target that is present in an image and the search background while the category clutter metric measured similarity by quantifying shared features between a categorical search target and the search background.

Using the data from the search experiment reported in Chapter 3, we compared the performance of these clutter metrics to a target-agnostic clutter metric in predicting search behavior.

Our results showed that both the exemplar clutter metric and the category clutter metric predicted search performance. In particular, in target present images, both metrics outperformed the target-agnostic metric. Our analysis using joint models suggested a significant contribution of the new clutter metrics beyond the target agnostic metric. Additionally, in target absent images, although the target-agnostic metric performed relatively better than the category clutter metric, adding the category clutter metric as a predictor to the model significantly improved its category clutter performance.

Most common objective clutter measures in psychophysics successfully predict performance in search tasks as a function of clutter (Bravo & Farid, 2008; Lohrenz et al., 2009; Mack & Oliva, 2004; Rosenholtz et al., 2005, 2007; Yu et al., 2014). However, these measures are agnostic to search targets and do not account for the target-background similarity in quantifying performance. Clutter measures in image optics consider the target-background similarity (Chang & Zhang, 2006; Chu et al., 2012; Moore et al., 2010a, 2010c; Schmieder & Weathersby, 1983; Silk, 1995; Tidhar et al., 1994); however, most measures do so at the exemplar level. As a result, these metrics can not determine performance in target absent cases. Few models which account for the target-background similarity also attempted to measure similarity for generic targets. Camp et al. (2013) constructed generic targets by segmenting instances of targets and superimposing these in the pixel space based on their physical dimensions. However, this definition of a generic target results in a broad frequency spectrum for a target representation leading to a loss of information. In our implementation of the category clutter metric, we represented categorical targets using a distribution of feature vectors, which led

to a richer representation of the target category.

A strong metric which accounts for similarity at the categorical level should be orientation and scale invariant. For a human observer, two instances of an object category that are positioned at two different orientations or captured at two different scales would be categorized in the same object category. Therefore, a good measure of similarity should consider these instances as similar to each other. For example, a set of keys which were oriented diagonally should be marked as similar to a set of keys which were oriented horizontally. Similarly, estimates of similarity should not vary as a function of different sizes or scales of the keys. Our categorical clutter metric used the actual orientation of objects as they are positioned in images and a single object size (i.e., median object size for a category). Therefore, implementing orientation and scale invariance can improve the performance of these metrics.

Additionally, a good measure of similarity should be independent of contrast. Two images of a set of keys with two different contrast levels would probably be considered as similar to each other by a human observer. The exemplar clutter metric uses a normalization procedure to eliminate the effect of contrast while measuring similarity while the category clutter metric does not use a normalization procedure.

Moreover, our metrics only use the information provided by the orientation subbands as the relevant feature to quantify similarity. Other features can also be considered. HOG feature detectors can be implemented to extract relevant features. We aim to remedy these issues in the future work.

## 4.5 Conclusion

We propose two new clutter metrics which account for the target-background similarity, both at the exemplar and category level, in predicting search performance. Our results demonstrate that both metrics predict search performance, suggesting that the target-background similarity should be incorporated into models of visual search and clutter in determining performance in natural search tasks.

## Chapter 5

### General Conclusions

Visual search is an essential part of everyday tasks. We constantly scan the visual environment to locate items or to navigate through it. While carrying out such tasks, our visual system has to account for uncertainty and variability within the environment. Performance in visual search tasks depends on several factors (see Eckstein, 2011, for a review).

Visual clutter is a significant factor which determines performance in visual search tasks. As the amount of clutter increases, search performance decreases (Bravo & Farid, 2004, 2008; Lohrenz et al., 2009; Rosenholtz et al., 2005, 2007; Yu et al., 2014). The first goal of this dissertation was to examine the effects of clutter on search performance while considering different sources of uncertainty.

A particular type of uncertainty that is fundamental to visual search is the position uncertainty which can arise from various sources. The *extrinsic* sources of position uncertainty relate to properties of the task (i.e., possible target locations or set size) while the *intrinsic* sources of position uncertainty relate to properties of the observer (i.e., limitations of the human visual system).

While studies on position uncertainty has commonly examined extrinsic sources (e.g., Bochud et al., 2004; Burgess & Ghandeharian, 1984b; Swensson & Judy, 1981), a few studies have particularly examined intrinsic sources (e.g., Michel & Geisler, 2011; Pelli, 1985; Tanner, 1961). In the context of visual searches, the detrimental effects of intrinsic position uncertainty can be investigated by manipulating the visual clutter in the search environment. Using an experimental

technique to manipulate synthetic clutter in visual displays, we showed that intrinsic position uncertainty impairs overt search performance in synthetic noise displays (Semizer & Michel, 2017). Whether these effects of intrinsic uncertainty generalize to searches in more naturalistic contexts is an open question, which we attempted to answer in this dissertation.

Target-background similarity is also a significant factor which determines performance in visual search tasks. Targets that share similar features to the background or to other items in the display impair the search performance (J. Duncan & Humphrey, 1989; Wolfe et al., 2002). However, most of the existing clutter measures in psychophysics do not account for features of search targets (Bravo & Farid, 2008; Lohrenz et al., 2009; Mack & Oliva, 2004; Rosenholtz et al., 2005, 2007; Yu et al., 2014) while the existing clutter measures in image optics account for some target features but they are limited and mostly not for categorical targets (Chang & Zhang, 2006; Chu et al., 2012; Moore et al., 2010a, 2010c; Schmieder & Weathersby, 1983; Silk, 1995; Tidhar et al., 1994). Thus, the second goal of this dissertation was to develop a clutter metric which takes features of search targets into account while predicting effects of clutter on search performance.

We designed two studies to investigate these questions and to achieve the aforementioned goals. In Study I, we examined the effect of clutter on search performance independently of set size. We quantified the existing clutter in natural scenes and measured the search performance when controlling the set size. Our results demonstrate that clutter impairs search performance, and does so independently of set size. These findings suggest that intrinsic position uncertainty limits search in natural scenes in the similar way it limits search in synthetic displays.

In Study II, we proposed two new clutter metrics which measure target-background similarity at the exemplar and categorical levels. The exemplar clutter metric made use of target features present in images and quantified similarity by comparing these features to background features. The category clutter metric made use of categorical target features and quantified similarity by comparing distribution of these features to background features. Both metrics used the orientation subbands as the relevant feature while also accounting for target size. Both clutter metrics predicted search performance, suggesting that the target-background similarity should be incorporated into models of visual search and clutter in determining performance in natural search tasks.

Among the significant contributions of this dissertation, effects of clutter on search performance in natural scenes were disambiguated from the effects of set size, intrinsic position uncertainty was proposed as a limiting factor in naturalistic searches, two new clutter metrics which account for target-background similarity were developed, and a new method of predicting search performance in the absence of a search target was introduced.

When combined, these two studies have implications for models of search and clutter. First, intrinsic position uncertainty should be incorporated into models which account for search performance. Previously, intrinsic position uncertainty was identified as a major limiting factor for detection and localization performance in single fixation search tasks (Michel & Geisler, 2011) as well as in multiple fixation search tasks (Semizer & Michel, 2017). Here we presented evidence that these detrimental effects of intrinsic position uncertainty persist even in more naturalistic searches.

Second, models of search and clutter incorporate measures of target-background

similarity while explaining search performance. We presented evidence that features of search targets and their interaction with the search background are important factors in quantifying performance in search tasks. In particular, accounting for the similarity between a search image and a search target category can improve the predictive power of search models.

This work contributed to the knowledge on position uncertainty, visual search, and clutter. Nonetheless, there are still open questions that need to be investigated in the context of visual search and clutter. In particular, potential implementations of the methods developed and presented in this dissertation require future work. The proposed clutter metrics can be improved by adding methods to achieve orientation and scale invariance. It would be interesting to incorporate other similarity features to proposed clutter metrics. Additionally, testing these clutter metrics with different set of images or target categories would help evaluate their performance. Moreover, other distance measures can be implemented to quantify similarity. Finally, performance of these metrics can be compared to other existing metrics from image-optics which also account for target-background similarity.

## References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csáki (Eds.), *2nd international symposium on information theory* (pp. 267–281). Budapest: Akadémiai Kiadó.
- Asher, M. F., Tolhurst, D. J., Troscianko, T., & Gilchrist, I. D. (2013). Regional effects of clutter on human target detection performance. *Journal of Vision*, *13*(5), 1–15. doi: 10.1167/13.5.25
- Baldassi, S., & Burr, D. C. (2000). Feature-based integration of orientation signals in visual search. *Vision Research*, *40*(10-12), 1293–1300. doi: 10.1016/S0042-6989(00)00029-8
- Barten, P. G. J. (1990). Integral Method. *JOSA A*, *7*(10), 2024–2031.
- Bochud, F. O., Abbey, C. K., & Eckstein, M. P. (2004). Search for lesions in mammograms: Statistical characterization of observer responses. *Medical Physics*, *31*(1), 24–36. doi: 10.1118/1.1630493
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, *226*, 177–178. doi: 10.1038/226177a0
- Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, *11*(5), 4. doi: 10.1167/11.5.4
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

- Bravo, M. J., & Farid, H. (2004). Search for a category target in clutter. *Perception*, 33(6), 643–652. doi: 10.1068/p5244
- Bravo, M. J., & Farid, H. (2008). A scale invariant measure of clutter. *Journal of Vision*, 8(2008), 23.1–9. doi: 10.1167/8.1.23
- Bravo, M. J., & Farid, H. (2009). The specificity of the search template. *Journal of Vision*, 9(1), 1–9. doi: 10.1167/9.1.34.Introduction
- Brettell, D. S., Berry, E., & Smith, M. A. (2007). The effect of experience on detectability in local area anatomical noise. *British Journal of Radiology*, 80(951), 186–193. doi: 10.1259/bjr/66715796
- Burgess, A. E. (1994). Statistically defined backgrounds: performance of a modified nonprewhitening observer model. *Journal of the Optical Society of America. A*, 11(4), 1237–42. doi: 10.1364/JOSAA.11.001237
- Burgess, A. E., & Ghandeharian, H. (1984a). Visual signal detection. I. Ability to use phase information. *Journal of the Optical Society of America. A*, 1(8), 900–905. doi: 10.1364/JOSAA.1.000900
- Burgess, A. E., & Ghandeharian, H. (1984b). Visual signal detection: II. Signal-location identification. *Journal of the Optical Society of America A: Optics and Image Science*, 1, 906–910. doi: 10.1364/JOSAA.1.000906
- Burgess, A. E., Li, X., & Abbey, C. K. (1997). Visual signal detectability with two noise components: anomalous masking effects. *Journal of the Optical Society of America. A*, 14(9), 2420–2442. doi: 10.1364/JOSAA.14.002420
- Cameron, E. L., Tai, J. C., & Carrasco, M. (2002). Covert attention affects the psychometric function of contrast sensitivity. *Vision Research*, 42(8), 949–967. doi: 10.1016/S0042-6989(02)00039-1
- Camp, H. A., Moyer, S., & Moore, R. K. (2010). Comparing masked target transform volume (MTTV) clutter metric to human observer evaluation of visual clutter. *Proceedings of SPIE*, 7662, 1–12. doi: 10.1117/12.851799

- Camp, H. A., Moyer, S., & Moore, R. K. (2013). Performance characteristics of the masked target transform volume clutter metric. *Optical Engineering*, 52(4), 041104–1–9. doi: 10.1117/1.OE.52.4.041104
- Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, 51(13), 1484–1525. doi: 10.1016/j.visres.2011.04.012
- Carrasco, M., Evert, D. L., Chang, I., & Katz, S. M. (1995). The eccentricity effect: Target eccentricity affects performance on conjunction searches. *Perception & Psychophysics*, 57(8), 1241–1261. doi: 10.3758/BF03208380
- Carrasco, M., Giordano, A. M., & McElree, B. (2004). Temporal performance fields: Visual and attentional factors. *Vision Research*, 44(12), 1351–1365. doi: 10.1016/j.visres.2003.11.026
- Carrasco, M., & Yeshurun, Y. (1998). The contribution of covert attention to the set-size and eccentricity effects in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 24(2), 673–692. doi: 10.1037/0096-1523.24.2.673
- Caspi, A., Beutter, B. R., & Eckstein, M. P. (2004). The time course of visual information accrual guiding eye movement decisions. *Proceedings of the National Academy of Sciences of the United States of America*, 101(35), 13086–90. doi: 10.1073/pnas.0305329101
- Castelhano, M. S., & Heaven, C. (2011). Scene context influences without scene gist: Eye movements guided by spatial associations in visual search. *Psychonomic Bulletin & Review*, 18(5), 890–896. doi: 10.3758/s13423-011-0107-8
- Castella, C., Eckstein, M. P., Abbey, C. K., Kinkel, K., Verdun, F. R., Saunders, R. S., . . . Bochud, F. O. (2009). Mass detection on mammograms: influence of signal shape uncertainty on human and model observers. *Journal of the Optical Society of America A*, 26(2), 425–436. doi: 10.1364/JOSAA.26.000425

- Chang, H., & Zhang, J. (2006). New metrics for clutter affecting human target acquisition. *IEEE Transactions on Aerospace and Electronic Systems*, 42(1), 361–368. doi: 10.1109/TAES.2006.1603429
- Chang, H., Zhang, J., Liu, X., Yang, C., & Li, Q. (2010). Color image clutter metrics for predicting human target acquisition performance. In *6th international conference wireless communications networking and mobile computing (wicom)* (pp. 1–4). New York, NY: IEEE. doi: 10.1109/WICOM.2010.5600622
- Chu, X.-q., Yang, C., & Qian, L. (2012). Contrast-sensitivity-function-based clutter metric. *Optical Engineering*, 51(6), 067003. doi: 10.1117/1.OE.51.6.067003
- Cohn, T. E., & Lasley, D. J. (1974). Detectability of a luminance increment: Effect of spatial uncertainty. *Journal of the Optical Society of America*, 64, 1715–1719. doi: 10.1364/JOSA.66.001426
- Cohn, T. E., & Wardlaw, J. C. (1985). Effect of large spatial uncertainty on foveal luminance increment detectability. *Journal of the Optical Society of America A: Optics and Image Science*, 2, 820–825. doi: 10.1364/JOSAA.2.000820
- De Valois, R. L., & De Valois, K. K. (1988). *Spatial Vision*. New York: Oxford University Press.
- Donovan, T., & Litchfield, D. (2013). Looking for Cancer: Expertise Related Differences in Searching and Decision Making. *Applied Cognitive Psychology*, 27(1), 43–49. doi: 10.1002/acp.2869
- Dosher, B. A., & Lu, Z. L. (2000a). Mechanisms of perceptual attention in precuing of location. In *Vision research* (Vol. 40, pp. 1269–1292). doi: 10.1016/S0042-6989(00)00019-5
- Dosher, B. A., & Lu, Z.-l. (2000b). Noise Exclusion in Spatial Attention. *Psychological Science*, 11(2), 139–146. doi: 10.1111/1467-9280.00229

- Drew, T., Evans, K., V $\tilde{o}$ , M. L. H., Jacobson, F. L., & Wolfe, J. M. (2013). Informatics in radiology: What can you see in a single glance and how might this guide visual search in medical images? *Radiographics*, *33*(1), 263–275. doi: 10.1148/rg.331125023
- Duncan, J., & Humphrey, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433–458.
- Duncan, R. O., & Boynton, G. M. (2003). Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron*, *38*(4), 659–671. doi: 10.1016/S0896-6273(03)00265-4
- Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision*, *11*(5), 14–14. doi: 10.1167/11.5.14
- Eckstein, M. P., Beutter, B. R., Pham, B. T., Shimozaki, S. S., & Stone, L. S. (2007). Similar Neural Representations of the Target for Saccades and Perception during Search. *Journal of Neuroscience*, *27*(6), 1266–1270. doi: 10.1038/nn1607
- Eckstein, M. P., Thomas, J. P., Palmer, J., & Shimozaki, S. S. (2000). A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Perception & Psychophysics*, *62*(3), 425–451. doi: 10.3758/BF03212096
- Eckstein, M. P., & Whiting, J. S. (1996). Visual signal detection in structured backgrounds. I. Effect of number of possible spatial locations and signal contrast. *Journal of the Optical Society of America. A*, *13*(9), 1777–1787. doi: 10.1364/JOSAA.13.001777
- Egeth, H., Atkinson, J., Gilmore, G., & Marcus, N. (1973). Factors affecting processing mode in visual search. *Perception & Psychophysics*, *13*(3), 394–402. doi: 10.3758/BF03205792
- Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a

- glance of a real-world scene? *Journal of vision*, 7(1), 10. doi: 10.1167/7.1.10
- Felzenszwalb, P. F., & Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2), 167–181. doi: 10.1023/B:VISI.0000022288.19776.77
- Findlay, J. M. (1997). Saccade target selection during visual search. *Vision Research*, 37(5), 617–631. doi: 10.1016/S0042-6989(96)00218-0
- Fischer, B. (1973). Overlap of receptive field centers and representation of the visual field in the cat's optic tract. *Vision Research*, 13(11), 2113–2120. doi: 10.1016/0042-6989(73)90188-0
- Fleck, M. S., & Mitroff, S. R. (2007). Rare targets are rarely missed in correctable search. *Psychological Science*, 18(11), 943–947. doi: 10.1111/j.1467-9280.2007.02006.x
- Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General*, 108(3), 316–355. doi: 10.1037/0096-3445.108.3.316
- Geisler, W. S., & Chou, K. L. (1995). Separation of low-level and high-level factors in complex tasks: visual search. *Psychological Review*, 102(2), 356–378. doi: 10.1037/0033-295X.102.2.356
- Geisler, W. S., Perry, J. S., & Najemnik, J. (2006). Visual search: The role of peripheral information measured using gaze-contingent displays. *Journal of Vision*, 6(9), 1. doi: 10.1167/6.9.1
- Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.
- Greene, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, 58(2), 137–176. doi: 10.1016/j.cogpsych.2008.06.001

- He, G., Zhang, J., Liu, D., & Chang, H. (2008). Clutter metric based on the Cramer-Rao lower bound on automatic target recognition. *Applied Optics*, 47(29), 5534–40. doi: 10.1364/AO.47.005534
- Henderson, J. M., Chanceaux, M., & Smith, T. J. (2009). The influence of clutter on real-world scene search : Evidence from search efficiency and eye movements. *Journal of Vision*, 9(1), 1–8. doi: 10.1167/9.1.32
- Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The Effects of Semantic Consistency on Eye Movements During Complex Scene Viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 25(1), 210–228. doi: 10.1037/0096-1523.25.1.210
- Hollingworth, A. (2006). Visual memory for natural scenes: Evidence from change detection and visual search. *Visual Cognition*, 14(4-8), 781–807. doi: 10.1080/13506280500193818
- Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. *Nature*, 394(6693), 575–577. doi: 10.1038/29068
- Kleene, N., & Michel, M. (2018). The capacity of trans-saccadic memory in visual search. *Psychological Review*, 125(3), 391–408. doi: 10.1037/rev0000099
- Klein, R. M. (1988). Inhibitory tagging system facilitates visual search. *Nature*, 334(6181), 430–431. doi: 10.1038/334430a0
- Klein, R. M., & Macinnes, W. J. (1999). Inhibition of return is a foraging facilitator in visual search. *Psychological Science*, 10(4), 346–352. doi: 10.1111/1467-9280.00166
- Kristjánsson, Á. (2000). In Search of Remembrance: Evidence for Memory in Visual Search. *Psychological Science*, 11(4), 328–332. doi: 10.1111/1467-9280.00265
- Krumhansl, C. L., & Thomas, E. A. C. (1977). Effect of level of confusability on reporting letters from briefly presented visual displays. *Perception &*

- Psychophysics*, 21(3), 269–279. doi: 10.3758/BF03214239
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79–86.
- Kunar, M. A., Rich, A. N., & Wolfe, J. M. (2010). Spatial and temporal separation fails to counteract the effects of low prevalence in visual search. *Visual Cognition*, 18(6), 881–897. doi: 10.1038/jid.2014.371
- Lesgold, A., Robinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing x-ray pictures. In *The nature of expertise* (pp. 311–342). Lawrence Erlbaum Associates, Inc.
- Levi, D. M. (2008). Crowding-An essential bottleneck for object recognition: A mini-review. *Vision Research*, 48(5), 635–654. doi: 10.1016/j.visres.2007.12.009
- Levi, D. M., Hariharan, S., & Klein, S. A. (2002). Suppressive and facilitatory spatial interactions in peripheral vision: Peripheral crowding is neither size invariant nor simple contrast masking. *Journal of vision*, 2(2), 3. doi: 10.1167/2.2.3
- Levi, D. M., Klein, S. A., & Aitsebaomo, A. P. (1985). Vernier acuity, crowding and cortical magnification. *Vision Research*, 25(7), 963–977. doi: 10.1016/0042-6989(85)90207-X
- Lohrenz, M. C., Trafton, J. G., Beck, M. R., & Gendron, M. L. (2009). A model of clutter for complex, multivariate geospatial displays. *Human Factors*, 51(1), 90–101. doi: 10.1177/0018720809333518
- Lu, Z. L., & Doshier, B. A. (1998). External noise distinguishes mechanisms of attention. In *Vision research* (Vol. 38, pp. 1183–1198). doi: 10.1016/B978-012375731-9/50078-1
- Lu, Z.-L., & Doshier, B. A. (2000). Spatial attention: different mechanisms for central and peripheral temporal precues? *Journal of Experimental*

- Psychology: Human Perception and Performance*, 26(5), 1534–48. doi: 10.1037//0096-1523.26.5.1534
- Luck, S. J., & Vogel, E. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657), 279–281. doi: 10.1038/36846
- Mack, M. L., & Oliva, A. (2004). Computational estimation of visual complexity. Poster presented at the Twelfth Annual Object, Perception, Attention, and Memory Conference, Minneapolis, MN.
- Mahalanobis, P. C. (1936). On the generalized distance in statistics. In *Proceedings of the national institute of sciences of india* (pp. 2:49–55).
- Manjeshwar, R. M., & Wilson, D. L. (2001). Hyperefficient detection of targets in noisy images. *Journal of the Optical Society of America. A, Optics, image science, and vision*, 18(3), 507–13.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). Academic Press.
- Michel, M., & Geisler, W. (2009). Gaze contingent displays: Analysis of saccadic plasticity in visual search. *Society for Information Display Technical Digest*, 40(1), 911–914. doi: 10.1889/1.3256945
- Michel, M., & Geisler, W. S. (2011). Intrinsic position uncertainty explains detection and localization performance in peripheral vision. *Journal of Vision*, 11(1), 1–18. doi: 10.1167/11.1.18
- Miller, G. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 101(2), 343–352. doi: 10.1037/h0043158
- Moore, R. K., Camp, H. A., Moyer, S., & Halford, C. E. (2010a). Masked target transform volume clutter metric applied to vehicle search. In *Proceedings*

- of spie* (Vol. 7662, pp. 76620M–1–11). doi: 10.1117/12.850429
- Moore, R. K., Camp, H. A., Moyer, S., & Halford, C. E. (2010b). Masked target transform volume clutter metric applied to vehicle search. In *Proceedings of spie* (Vol. 7662). doi: 10.1117/12.850429
- Moore, R. K., Camp, H. A., Moyer, S., & Halford, C. E. (2010c). Triangle search experiment to isolate scene clutter effects. In *Proceedings of spie* (pp. 76620C–1–9). doi: 10.1117/12.850457
- Murray, R. F., Beutter, B. R., Eckstein, M. P., & Stone, L. S. (2003). Saccadic and perceptual performance in visual search tasks. II. Letter discrimination. *Journal of the Optical Society of America. A, Optics, image science, and vision*, 20(7), 1356–70. doi: 10.1364/JOSAA.20.001356
- Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, 434, 387–391. doi: 10.1167/5.8.778
- Najemnik, J., & Geisler, W. S. (2008). Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision*, 8(3), 4.1–14. doi: 10.1167/8.3.4
- Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, 46(5), 614–621. doi: 10.1016/j.visres.2005.08.025
- Neider, M. B., & Zelinsky, G. J. (2008). Exploring set size effects in scenes: Identifying the objects of search. *Visual Cognition*, 16(1), 1–10. doi: 10.1080/13506280701381691
- Neider, M. B., & Zelinsky, G. J. (2011). Cutting through the clutter: searching for targets in evolving complex scenes. *Journal of Vision*, 11(14), 1–16. doi: 10.1167/11.14.7
- Oliva, A., & Torralba, A. (2006). Building the gist of a scene: the role of global image features in recognition. *Progress in Brain Research*, 155, 23–36. doi:

10.1016/S0079-6123(06)55002-2

- Palmer, J. (1994). Set-size effects in visual search: The effect of attention is independent of the stimulus for simple tasks. *Vision Research*, 34(13), 1703–1721. doi: 10.1016/0042-6989(94)90128-7
- Palmer, J. (1995). Attention in visual search: Distinguishing four causes of a set-size effect. *Current Directions in Psychological Science*, 4(4), 118–123. doi: 10.1111/1467-8721.ep10772534
- Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research*, 40(10-12), 1227–1268. doi: 10.1016/S0042-6989(99)00244-8
- Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A*, 2, 1508–1532. doi: 10.1364/JOSAA.2.001508
- Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision*, 4(12), 1136–1169. doi: 10.1167/4.12.12
- Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *Journal of Vision*, 7(2), 20. doi: 10.1167/7.2.20
- Popple, A. V., & Levi, D. M. (2005). The perception of spatial order at a glance. *Vision Research*, 45(9), 1085–1090. doi: 10.1016/j.visres.2004.11.008
- Potter, M. C. (1976). Short-Term Conceptual Memory for Pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 2(5), 509–522. doi: 10.1037/0278-7393.2.5.509
- Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Experimental Psychology*, 81(1), 10–15. doi: 10.1037/h0027470
- Rajashekar, U., Bovik, A. C., & Cormack, L. K. (2006). Visual search in noise:

- revealing the influence of structural cues by gaze-contingent classification image analysis. *Journal of Vision*, 6(4), 379–86. doi: 10.1167/6.4.7
- Reingold, E. M., & Sheridan, H. (2011). Eye movements and visual expertise in chess and medicine. In (pp. 528–550).
- Rich, A. N., Kunar, M. A., Van Wert, M. J., Hidalgo-Sotelo, B., Horowitz, T. S., & Wolfe, J. M. (2008). Why do we miss rare targets? Exploring the boundaries of the low prevalence effect. *Journal of Vision*, 8(15), 1–17. doi: 10.1167/8.15.15
- Rosenholtz, R. (1999). Rapid communication A simple saliency model predicts a number of motion popout phenomena. *Vision Research*, 39, 3157–3163.
- Rosenholtz, R., Huang, J., Raj, A., Balas, B., & Ilie, L. (2012). A summary statistic representation in peripheral vision explains visual search. *Journal of Vision*, 12(4), 14–14. doi: 10.1167/12.4.14
- Rosenholtz, R., Li, Y., Mansfield, J., & Jin, Z. (2005). Feature congestion, a measure of display clutter. In *Sigchi* (pp. 761–770). Portland, Oregon.
- Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, 7(2), 17.1–22. doi: 10.1167/7.2.17
- Schmieder, D. E., & Weathersby, M. R. (1983). Detection Performance in Clutter with Variable Resolution. *IEEE Transactions on Aerospace and Electronic Systems*, AES-19(4), 622–630. doi: 10.1109/TAES.1983.309351
- Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time-and spatial-scale-dependent scene recognition. *Psychological Science*, 5(4), 195–200.
- Semizer, Y., & Michel, M. (2017). Intrinsic position uncertainty impairs overt search performance. *Journal of Vision*, 17(9), 1–17. doi: 10.1167/17.9.13
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27, 623–656. doi: <http://dx.doi.org/10.1145/>

584091.584093

- Silk, J. (1995). Statistical variance analysis of clutter scenes and applications to a target acquisition test. *P-2950, Institute for Defense Analyses*.
- Simoncelli, E., & Freeman, W. (1995). The steerable pyramid: A flexible architecture for multi-scale derivative computation. In *Proceedings of the 2nd ieee international conference on image processing* (Vol. 3, pp. 444–447). Washington, DC.. doi: 10.1109/ICIP.1995.537667
- Swensson, R. G., & Judy, P. F. (1981). Detection of noisy visual targets: models for the effects of spatial uncertainty and signal-to-noise ratio. *Perception & Psychophysics*, 29(6), 521–534. doi: 10.3758/BF03207369
- Tanner, W. P. (1961). Physiological implications of psychophysical data. *Annals of the New York Academy of Sciences*, 89, 752–765. doi: 10.1111/j.1749-6632.1961.tb20176.x
- Tidhar, G., Reiter, G., Avital, Z., & Hadar, Y. (1994). Modeling human search and target acquisition performance: IV. Detection probability in the cluttered environment. *Optical Engineering*, 33(3), 801–808.
- Torralba, A., Oliva, A., Castelano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological Review*, 113(4), 766–786. doi: 10.1037/0033-295X.113.4.766
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97–136. doi: 10.1016/0010-0285(80)90005-5
- van den Berg, R., Cornelissen, F. W., & Roerdink, J. B. T. M. (2009). A crowding model of visual clutter. *Journal of Vision*, 9(4), 24.1–11. doi: 10.1167/9.4.24
- Verghese, P. (2001). Visual search and attention: A signal detection theory approach. *Neuron*, 31(4), 523–535. doi: 10.1016/S0896-6273(01)00392-0

- Verghese, P. (2012). Active search for multiple targets is inefficient. *Vision Research*, *74*, 61–71. doi: 10.1016/j.visres.2012.08.008
- Vickery, T. J., King, L.-W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision*, *5*(1), 81–92. doi: 10.1167/5.1.8
- Williams, L. G. (1967). The effects of target specification on objects fixated during visual search. *Acta Psychologica*, *27*, 355–360. doi: 10.3758/BF03215795
- Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238. doi: 10.3758/BF03200774
- Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13–71). London: University College London Press.
- Wolfe, J. M., Brunelli, D. N., Rubinstein, J., & Horowitz, T. S. (2013). Prevalence effects in newly trained airport checkpoint screeners: trained observers miss rare targets, too. *Journal of Vision*, *13*(3), 33. doi: 10.1167/13.3.33
- Wolfe, J. M., Horowitz, T., & Kenner, N. (2005). Rare items often missed in visual searches. *Nature*, *435*(May), 6–7. doi: 435439a
- Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, *5*(6), 495–501. doi: 10.1038/nrn1411
- Wolfe, J. M., Horowitz, T. S., Van Wert, M. J., Kenner, N. M., Place, S. S., & Kibbi, N. (2007). Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of Experimental Psychology General*, *136*(4), 623–38. doi: 10.1037/0096-3445.136.4.623
- Wolfe, J. M., Oliva, A., Horowitz, T. S., Butcher, S. J., & Bompas, A. (2002). Segmentation of objects from backgrounds in visual search tasks. *Vision Research*, *42*(28), 2985–3004. doi: 10.1016/S0042-6989(02)00388-7
- Wolfe, J. M., & Van Wert, M. J. (2010). Varying Target Prevalence Reveals Two Dissociable Decision Criteria in Visual Search. *Current Biology*, *20*(2),

- 121–124. doi: 10.1016/j.cub.2009.11.066
- Wolfe, J. M., Võ, M. L. H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, 15(2), 77–84. doi: 10.1016/j.tics.2010.12.001
- Wolford, G. (1975). Perturbation model for letter identification. *Psychological Review*, 82(3), 184–199. doi: 10.1037/0033-295X.82.3.184
- Woodman, G. F., & Chun, M. M. (2006). The role of working memory and long-term memory in visual search. *Visual Cognition*, 14(4-8), 808–830. doi: 10.1080/13506280500197397
- Yu, C.-P., Samaras, D., & Zelinsky, G. J. (2014). Modeling visual clutter perception using proto-object segmentation. *Journal of Vision*, 14(7), 1–16. doi: 10.1167/14.7.4
- Zelinsky, G. J. (2008). A Theory of Eye Movements During Target Acquisition. *Psychological Review*, 115(4), 787–835. doi: 10.1037/a0013118
- Zhang, S., & Eckstein, M. P. (2010). Evolution and optimality of similar neural mechanisms for perception and action during search. *PLoS Computational Biology*, 6(9). doi: 10.1371/journal.pcbi.1000930
- Zhang, Y., Abbey, C. K., & Eckstein, M. P. (2006). Adaptive detection mechanisms in globally statistically nonstationary-oriented noise. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, 23(7), 1549–1558. doi: 90370[pil]
- Zhang, Y., Pham, B. T., & Eckstein, M. P. (2005). Task-based model/human observer evaluation of SPIHT wavelet compression with human visual system-based quantization. *Academic Radiology*, 12(3), 324–336. doi: 10.1016/j.acra.2004.09.015