The Pennsylvania State University

The Graduate School

College of Information Sciences and Technology

TASKING THE SOFT SENSOR: USING CUES TO DIRECT VISUAL ATTENTION

A Thesis in

Information Sciences and Technology

by

Alice Shapiro

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

August 2009

The thesis of Alice Shapiro was reviewed and approved* by the following:

David Hall,
Associate Professor of Information Sciences and Technology
Thesis Advisor

Michael McNeese
Associate Professor of Information Sciences and Technology

Frederico Fonseca
Associate Professor of Information Sciences and Technology

John Yen
Professor of Information Sciences and Technology
Associate Dean for Research and Graduate Programs of Information Science and Technology

*Signatures are on file in the Graduate School

Abstract

Maintaining awareness in an ever-changing environment is a difficult task. The observer's awareness at any given point in time is significantly impacted by the context and the goals of the observer. Despite the best intentions of the observer, there are a number of aspects of visual perception that are likely to cloud the observational accuracy of an individual performing a visual search task. In this experiment, I test the effect of a tool to allocate spatial attention in a movie clip of naturalistic video footage in an environment familiar to the observer. Observational awareness in viewing surveillance footage based on experience and expertise and the potential for technological assessment and improvement of visual search ability is discussed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Acknowledgement

Special thanks to Qian Xu, who was an excellent sounding board for my methods questions, and Ben Hellar and Bernie Zhang who were instrumental in the editing process.

Chapter 1

Introduction

From the perspective of digital information management, there are a number of different ways to filter, or search information for relevance. Data indexing and algorithmic isolation of objects on video surveillance are just a few. Information management from a human perspective is more complex and requires consideration of the cognitive limitations of human observers as soft sensors (Hall, Llinas, McNeese, and Mullen 2008).

The potential benefit of more effectively tasking humans as soft sensors is particularly pertinent given the increase in potential surveillance methods, and their increasing availability through networking. One of the terms that has been used to describe this situation is “Network Centric” (Wesensten, Belenky and Balkin 2005). With an increase in information sharing between individuals, filtering and search techniques are increasingly important. Network Centric Cognition has implications for information sharing between individuals in groups as well as the way that information is accessed by the individual observer (Wesensten et al. 2005). For the purpose of this experiment, I am interested in information gathering at primarily the individual level.

In order to accurately report events, an observer must be aware of important information in a given situation. The Army Field Manual defines Situational Awareness as

“*Knowledge and understanding of the current situation which promotes timely, relevant and accurate assessment of friendly, competitive and other operations within the battlespace in order to facilitate decision making. An informational perspective and skill that fosters an ability to determine quickly the context and relevance of events that are unfolding.*” Army Field Manual 1-02 (September 2004)

The Army Field manual definition effectively captures the need of the observer for contextual awareness and a continuing search of the environment. At the individual level, the observer faces a number of challenges when asked to observe, determine relevant details about a given situation, and accurately report those details. Therefore, the first steps in that process of gaining situational awareness from a visual perspective are identifying the context, correctly allocating attention to regions that are likely to be of interest, and searching the environment for people or objects of interest.

Observers may use a number of different heuristics in order to determine the best region in space to allocate their attention. Some common methods and goals observers use to conduct visual search will be discussed. This thesis addresses the question: Using the information we already know about visual search, how can we best direct observers to allocate their attention during the process of visual search?

1.1 Network Centric Cognition - Context

Network Centric Operations focus on the importance of all elements within a system in achieving a state of total situational awareness, where all relevant information is shared between individuals in a timely and accurate manner (Wesensten et al. 2005). Through successful attentional allocation, the hope is that humans acting as soft sensors will be able to more rapidly attain situational awareness and relay their findings accurately.

Total situational awareness is dependent on the observer’s ability to accurately report events, but is also dependent on interpretation and direction of entities with a larger knowledge base. Through more accurate attentional allocation, these entities, both technological and human, could share their specific knowledge of a situation to theoretically help allocate the attention of observers (See Figure 1-1). Currently, this form of attentional direction occurs through radio and other forms of traditional communication (Pravia et al. 2008). Verbal goal direction, however, is often ambiguous and inefficient, particularly in the context of surveillance video (Ferrin et al. 2008). One of the goals of attentional allocation is to more efficiently allow information transfer to the level of the observer.

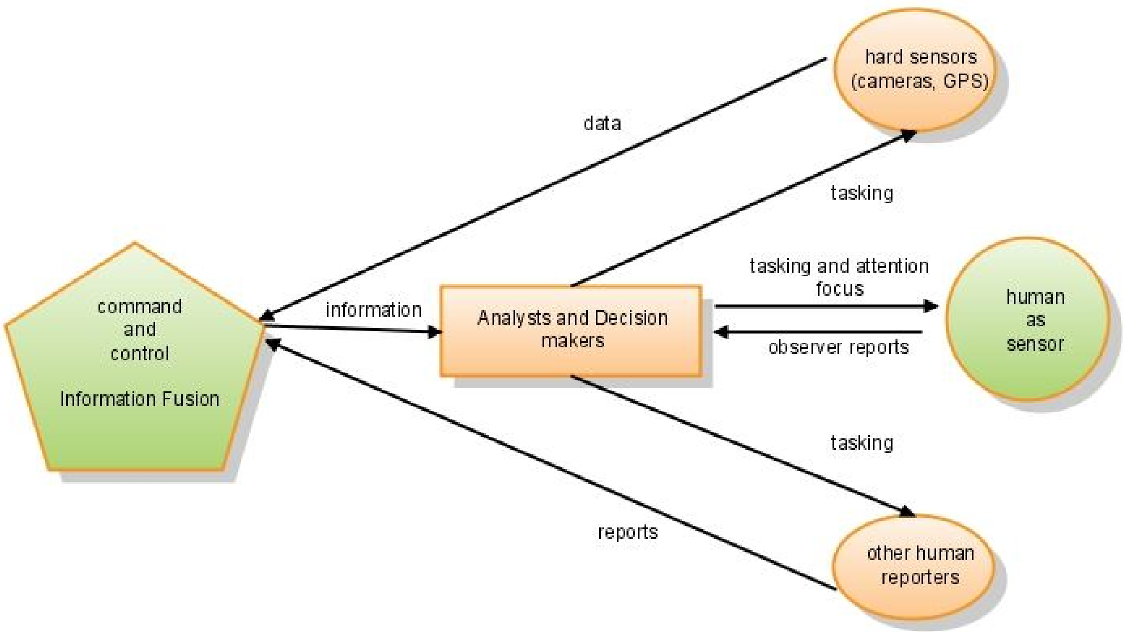


Figure 1-1. Information flow to and from humans as soft sensors.

1.2 Purpose and Intents (Motivations)

While there are many definitions of situational awareness that expand on the Army Field Manual's (2004) definition, I focus on Endsley's definition which outlines the different levels of Situational Awareness as Perception, Comprehension, and Projection (2000). The aim of this thesis is to examine one way to improve the situational awareness of humans as soft sensors by allocating their visual attention with an attention directing tool.

There has recently been a call specifically for the combination of human feedback (soft) data to combine with surveillance data (Ferrin et al. 2008). This focuses on the need for the fusion of hard, movement and speech related data from automated surveillance sources with soft interpretations by human observers who can combine the two sources and give a less ambiguous interpretation. For example, humans as soft sensors can act to direct hard sensors to specific locations; one of the goals of hard and soft information fusion is to pool information from hard and soft sensors to track targets (Pravia 2009). Given the importance of visual awareness and search in the process of target acquisition and tracking, the role of human soft sensors as observers will be considered for the purpose of this thesis.

Despite the fact that humans as soft sensors generally provide less reliable information regarding quantitative constructs such as location and speed (Pravia 2009), Lossau and Stotz (2008) note that humans as soft sensors have a number of advantages over automated systems when it comes to the process of information acquisition and comprehension. This includes an ability to understand unstructured data, an ability to search out necessary data, process “bad data”, and understand information within a context. For surveillance information, all of these advantages are important. Information from any display source is comprehensible without conversion into a specific format, humans as soft sensors can turn to external sources for assistance if they are unable to understand an element of a situation, and they may be able to pick important pieces out of low quality video footage.

Perhaps the most important advantage for humans as soft sensors in a surveillance situation is the ability to understand events and objects in context. As Ferrin et al. (2008) state, humans as soft sensors can combine elements within a scene, but they can also draw on previous knowledge of situations to enable a better understanding. Scott-Brown and Cronin (2008) note that while there is a great body of work on visual cognition, this work is not always applied in the human interface design of surveillance systems, or in the general environmental designs of the workplace where these systems exist. This thesis aims to examine a few of the many ways in which visual cognition may affect surveillance for security purposes.

1.3 Summary of Accomplishments

A summary of the accomplishments in this thesis include:

- A literature review
- Experimental design construction
- Design and implementation of an attention-directing instrument
- Participant recruitment and experimental conduct
- An analysis of the experimental results
- A set of recommendations for future research

The first step in this project was to conduct a literature review of the areas relevant to the thesis topic. The literature review covers the difficulties encountered in the visual search of a changing visual scene as well as some of the specific issues relevant to the search of a naturalistic environment. Within this chapter, a few of the methods that observers employ in order to conduct visual search are discussed, and why the allocation of spatial attention is particularly relevant within the context of naturalistic image search. Some of the ways in which people develop expertise for a given task and how that expertise either does or does not carry over into visual search within different contexts are also discussed.

After a survey of the literature, an experiment designed to test the effects of spatial attention allocation on visual search in naturalistic video was designed and implemented.

Participants were recruited from both undergraduate and graduate sources. Verbal as well as email solicitation was employed within 3 different classrooms as well as two IST (Information Science and Technology) and SRA (Security and Risk Analysis) relevant list-servs. After running a practice session, participants were verbally directed to the target. Conditions included cued and uncued video footage.

Given the apparent importance of spatial cueing in the literature, it was necessary to design an experimental instrument in order to test the effects of spatial cueing. The experimental instrument was developed in several parts. The first step was acquiring appropriate video footage. At that point, several possible video clips with search targets were selected and pre-tested for difficulty. Appropriate video software was used to edit these clips, and to add cued locations. Macromedia Flash 8 was then used to create a user interface to pause the video, select targets, and send the response time information to a database for storage.

Chapter 2

Literature Review

2.1 Visual Search

For the purpose of this thesis, visual search is defined as the process of isolating unusual or goal directed items in a scene. Mobile devices and automated systems are increasingly used to aid observers in visual cognition tasks, such as wayfinding, using contextual information derived from hard sensors in a network centric environments (Xu et al. 2008, Want and Schilit 2001). Contextual awareness can be used as a filter to alter the types of information given to an observer based on their location, time of day, or other types of contextual elements (Xu et al. 2008).

Location aware computing, or computing where devices utilize network technologies such as general packet radio service to determine location, can be seen as a subset of context aware computing (Want and Schilit 2001). Dey (2001) suggests that there are three main functions that context aware computing supports: presentation of information and services, automatic service execution, and tagging for later retrieval. Software used to note a spatial location of interest would fit under the category of tagging for later retrieval. Many consumer services support cataloging of still images, either manually or through image recognition technology (facebook 2009, Apple 2009), supplementing human search and retrieval of images. An increasing number of applications also include some form of visual representation of what users see at eye level in an environment, supplementing visual search in natural environments (google 2009, Xu et al. 2008).This is true for a wide variety of contexts including museum guides, tourism guides, search engines, and other tools used in the guidance of individuals in unfamiliar environments (Xu et al. 2008). Davies et al. (2005) describe the use of a tour guide that combines information provided based on location awareness with an experimenter who acted “behind the scenes” to return information on images taken with a camera attached to a PDA, demonstrating the possible remote processing of imagery in a network centric environment. Interestingly, about half of the participants preferred to use the camera feature, while about half preferred to browse for information based on their location, indicated a possible individual difference in information seeking behavior in this type of environment.

In addition to technologies to tag and automatically process still images, a number of technologies are currently being developed to isolate important elements from surveillance video (Dee and Velastin 2008). Automated video detection systems often use statistical probabilities based on image identification and motion to make assessments of a given situation (Makris and Ellis 2002). Some of these methods include path analysis, which examines the usual route that people in a given environment take, and then notes those individuals who deviated from the path (Makris and Ellis 2002). Others use a more complex model to examine the probable goal related activity of multiple objects within the visual field based on their interactions (Intille and Bobick 2001). Currently, some of the applied automated detection systems make use of text based information for identification purposes such as license plate numbers (Hi-Tech Solutions 2007). Optical character recognition allows these systems to offer less ambiguity in their interpretation of visual information than systems based on multiple categories of image recognition.

One of the current suggestions for emphasizing information in video footage includes subtracting background or unimportant information entirely. This places the correct emphasis on a given category of people (those stopping in a specific location, traffic heading in a given direction) (Senior et al. 2005). This technique allows for effective attention allocation, but necessarily subtracts information from a scene, important when taking into account privacy concerns, as Senior et al. (2005) suggest, but potentially detrimental when little is known about the activity of a target.

The presentation of video information in combination with automated image recognition information presents another set of problems. While OCR information gleaned from license plate numbers is relatively straight forward to represent to an observer in text format, other types of information, such as moving objects in a scene, require a more graphical approach. One surveillance video filtering technique, utilized by Gorodnichy (2006), is to use automated recognition software to create a text file of events in a surveillance tape. This annotation is then supplemented with single frames containing high resolution images of faces, and colored, rectangular frames surrounding moving images on video. This particular software is particularly geared toward presenting information to a human observer after the fact. Other types of surveillance technologies have been designed to transmit real time streaming video both to humans and to automated recognition systems using limited bandwidth, as might be experienced using a mobile device (Gualdi, Prati, and Cucchiara 2008). Gualdi et al. (2008) note that the optimal video steam for human and for automated software may not be the same, as fluidity and the ability to rewind and replay is more important to human observers than for automated recognition systems.

The question addressed in this review is how best to allocate human attention to specific elements of a scene determined either by outside individuals or by automated software, while ignoring unimportant information. The primary research question with regards to visual search concerns whether or not spatial location cues effectively reduce search times. Previous research generally supports the idea that observers utilize spatial position information as a search strategy (Chun and Nakayama 2000). Observers tend to allocate attention based on previous knowledge of important areas within that scene, with feature and identifying figural information used to determine the most likely areas of interest. A spatial cue would therefore be one way to compensate for a lack of expertise or identifying feature and figural information within a scene, or a complex or crowded scene.

2.2 Why are cues necessary?

Some of the existing work in visual perception relating to naturalistic video footage is in the field of change detection (Simons and Chabris 1998, Simons et al. 2000). Visual search within the context of change blindness and inattentional blindness involves the comparison between two scenes, one before and one after a change has occurred, or general alertness to unusual changes in a motion picture respectively (Simons et al. 2000). Based on this literature, changes in a motion picture are not always immediately apparent, and are strongly dependent on the goals of the observer and their attentional allocation (Simons and Chabris 1998). In one study showing inattentional blindness, participants instructed to count basketball passes in a movie clip failed to notice a student in a gorilla costume walking through the scene (Simons and Chabris 1998, video at http://viscog.beckman.illinois.edu/flashmovie/15.php).

The features that individuals do notice in a scene are contextual. While participants were likely to notice the replacement of a student they were speaking to after their line of vision was interrupted, they were much less likely to notice the substitution of a different construction worker (Simons and Levin 1998). This example also illustrates the importance of visual continuity. Briefly covering an area of view during a change negatively impacts change detection, though change blindness may still occur without any occlusions (Simons, Franconeri, and Reimer 2000).

2.2.1 Distraction

It is not necessary for the entire field of view to be interrupted in order to hide a change, or even the part undergoing the change. Partial occlusion of a spatial area is sometimes referred to as a “mudsplash”, representing an irregularly shaped mask presented over part of a scene when viewed on a computer monitor or other display (O’Regan, Rensink, and Clark 1999). This can be seen as a stand in for a naturalistic situation where an object briefly occludes a given area in a scene. The mudsplashes do not need to occlude the target. The occlusions can therefore be thought of as spatial attention distracters. If only a few changes to a visual scene can be noticed at a given time, additional changes may overload the observer. The temporal relationship between changes also may overload an observer. Distracting changes that happen in rapid succession also tend to overload the observer, leading to change blindness (Raymond, Shapiro, and Arnell 1992).

In naturalistic situations, change or inattentional blindness may be one of the reasons why visual distractions are particularly problematic for individuals operating motor vehicles. Various changes to a visual scene that do not involve continuous motion, such as a traffic signal, or brake lights, may be readily detected at onset, but are more likely to escape detection if the driver is distracted during a change (National Safety

Commission 2007). Any salient distracters within the visual field would also tend to increase the chances of change blindness (Theuwees 2004).

2.2.2 Limited area of space for attention

The maximum area of space that an observer is capable of attending to is correlated with the ability to detect objects. Observers with a limited field of view are more likely to fall prey to inattentional or change blindness. One of the primary factors associated with traffic accidents in the elderly is a loss of effective field of view, which makes it less likely that elderly observers will notice changes on the road in the periphery of their vision (Owsley et al. 1998). Caird et al. (2005) directly tested the ability of elderly drivers in simulated driving scenarios to detect changes between one situation and another when only one element of a scene was changed, and a blank screen was presented between the two scenes. The performance was significantly worse for pedestrians that had entered intersections and for changes to street signs. Notably, expertise in a given context, such as chess, increases the area of possible change detection, pointing again to the importance of spatial attention in visual search (Reingold et al. 2001).

2.3 Basic issues in visual search

2.3.1 Expertise effects

Long term memory effects for a target in a given spatial location can develop rapidly and have been observed under controlled conditions. Given a series of similar backgrounds, observers tend to develop an implicit memory for the types of objects that will occur in similar spatial locations (Chun and Nakayama 2000). Another of the manifestations of learned spatial position is inhibition of return, a well established effect where observers have greater difficulty locating objects or locations they were previously instructed to ignore, evidencing the long term effects of goal directed search (Beck et al. 2006).

There is increasing evidence that observers also accrue useful long term memory information about the spatial location of objects in naturalistic contexts. For example, Myles-Worsley, Johnston, and Simons (1988) found that as Radiologists gained experience, the more likely they were to be able to spot pathological X-rays, but the less likely they were to correctly recognize and discriminate between normal X-rays. It was as though they began to ignore irrelevant information in the X-rays. An observer's learned ability to discern spatial locations of interest and distinguishing may only help them if the task is relevant to the learned and goal related .

Most expertise in naturalistic situations is highly tied to context. Werner and Thies (2000) found that although football aficionados were better able to notice changes in player configurations between one scene and another, they were not significantly better at noticing changes to the crowds in the stands. Reingold et al. (2001) found similar results for chess players who were better able than novices to detect changes in a chess piece presented in a possible game configuration, but not in a random configuration. The question remains whether experts in viewing every day events, such as police officers, are better able to perform search tasks than novice observers. Previous research suggests that novice observers are just as capable of detecting problematic human behaviors through the use of closed circuit television devices as experienced security personnel (Troscianko et al. 2002), however this was a type of search without a target specified on any given feature.

2.3.2 Novice experience effects

In order for spatial knowledge to positively impact response times, it needs to be very rapid. Rousselet, Joubert, and Fabre-Thorpe (2005) found that participants were correctly able to determine whether a natural scene depicted a target category or not with a 26 millisecond exposure. Accuracy was above 95%. Photos corresponded to the sea, to mountains, indoor environments, and urban environments. In general, the contents of naturalistic images tend to be processed more quickly than other types of images. Tatler and Melcher (2007) found that participants remembered the contents of naturalistic photographs more accurately than non-photographs, with the worst rates of position memory for non-photographic, non-realistic images. This indicates that even for novice observers, context can be used as a cue to spatial location.

More direct evidence to suggest that observers use spatial position to guide goal directed visual search comes from eye tracking data. In general, observers use eye movements to follow goal directed activities. Early evidence comes from Yarbus (1967), who documented directed eye movements over a photograph to important areas of a picture when observers were asked to determine the wealth of a family. A similar study by Torralba et al. (2006) found that the type of target can have a strong effect on the direction of gaze, even in the absence of a target. Torralba et al. (2006), for example, found that observers were more likely to look at a coffee table for a mug, and at a wall for a painting, with or without a painting in the scene, suggesting that observers use familiarity with different types of scenes to guide their visual gaze.

One of the recurrent patterns of visual scanning behavior appears to be multiple returns to the same areas of a scene, which lend credence to the claim that very little of a visual scene is actually stored in memory, although the spatial locations of objects may be remembered and searched out for later clarification (Chun 2000, Peterson et al. 2001). Hoffman and Subramanium (1995) found a strong relationship between the direction of a saccade (brief eye movement) and visual attention. Targets placed near the location of a saccade were significantly more likely to be noticed than targets placed spatially distant from a saccade.

Research on change blindness has also found that if a scene is altered during a saccade, participants are much less likely to notice (Aivar et al. 2005). Aivar et al. (2005) explain this as form of goal directed gaze similar to that found by Hayhoe (2000). As one might suspect, observers direct their gaze to objects involved in carrying out an activity. Individuals carrying out a copying task (manually replicating a line drawing) accurately remembered the spatial location they had just viewed, even though location of an object had been moved (Aivar et al. 2005). This finding points to good memory for spatial location, and the fact that observers do not have to continually re-scan the environment between saccades. If an object changes, or is occluded between saccades, there will be a time cost to find that object again. Spatial location is used as a cue to carry out visual tasks, while features are not, therefore pointers to spatial location may be faster than feature based cues.

2.4 Effects of Complexity on Image Recognition

Some of the research on reaction time tools in a Network Centric context does not involve recognition of complex objects. The Psychomotor Vigilance Test (PVT), for example, represents a timed response to a literal target symbol on a palm pilot, and is discussed more in terms of task-switching performance rather than as a tool to respond to external visual stimuli (Thorne et al. 2005). This limits the ecological validity of the PVT. The PVT tool is more effectively designed to measure speed of information reporting or response time in general rather than information comprehension.

It takes time to view an object and recognize it. Search of naturalistic images takes longer than search of simple shapes. Current theories of visual memory suggest that the representation of a scene is impoverished and limited by short term memory (Simons and Levin 1998). Research within the change blindness paradigm seems to suggest that time is required to successfully encode complex stimuli. For simple shapes the required time is less than a second (Brady et al. 2008). The motion picture, and vision in general, represent a constantly shifting image. While the time duration necessary to encode an object in enough detail to accurately notice a later change is approx. 200 ms per item or less (Moore, Egeth Berglan and Luck 1996), Brady et al. (2008) found that for a collection of 6 items, 18 seconds of viewing (3 second per item) was required to notice minor changes in the object approximately 90% of the time. Time course is also a factor. Items more recently attended to (as measured by eye movement) are more likely to be remembered (Irwin and Zelinsky 2002). Irwin and Zelinsky (2002) reported that the location and identity of a maximum of 5 objects in a naturalistic scene was accurately remembered by observers, however, observers were only able to view a scene for a maximum of 5 seconds, which may have been the delimiting factor in object recognition.

Naturalistic visual search may engage a different type of search strategy than those that occur in artificial environments. Visual search is a goal oriented process, and it would seem to logically follow that individuals would use the presence of salient stimuli to engage in visual search. Salience is a term generally applied to colors and simple shapes that can be visually isolated with ease from their surroundings (Theeuwes 2004). There is some discussion as to whether items that are unique on a salient feature, referred to as singletons, can ever be completely ignored, or whether their presence automatically attracts attention, increasing response time (Theeuwes 2004). Bacon and Egeth (1994) contend that they can be ignored given the appropriate goal. Certainly, they can be isolated from their surroundings with relative ease. There is some evidence that given enough spatial distance between two stimuli of different colors, neurons at the higher levels of visual cortex selectively fail to fire (Moran and Desimone1985). This also suggests that with appropriate goal direction, salient singletons can be ignored. In naturalistic settings, observers may be forced to use less salient cues due to the complexity of input. This makes spatial attention and figure recognition more important. It is much less likely that a search target will be unique on a single feature such as color or orientation in a naturalistic setting than in an artificial environment (Williams et al. 2005). Scott-Brown and Cronin (2008) also cast doubts over whether color in CCTV footage used for security purposes would be especially useful, considering that lighting conditions are often poor and influenced by sodium lights, or other types of lights that do not emit the full spectrum, creating a false sense of color. Maloney and Wandell (1992), however, suggest that color constancy, or the ability to perceive color as invariant regardless of different lighting conditions, varies depending on the exact lighting conditions and can be improved with observer knowledge of the conditions.

2.5 Search strategies

The search for stimuli that are unique on a given feature is generally referred to as parallel search, since it is not dependent on the number of items in a display (Egeth and

Yantis 1997). In the case of a complex target that is not unique on any one feature, serial search is required. In this type of search, the features of each item in conjunction with each other must be considered separately. Serial search times increase systematically with the number of items in a given scene (Luck and Vogel 1997, Duncan and Humphreys 1989). There is some discussion in the literature on visual search as to whether search times increase systematically due to consideration of each image separately, or whether features within a scene are considered with increasing complexity over time (Egeth and Dagenbach 1991), but in either case, searching for a target that is not unique for any one salient feature increases response time.

2.5.1 Search Strategies for Naturalistic Images

When confronted with naturalistic images, previous studies seem to indicate that individuals are biased more towards allocating attention to irrelevant, but semantically similar stimuli than to irrelevant but visually salient (similarly colored) stimuli (Williams, Henderson and Zacks 2005). Despite the fact that color is a highly salient cue (Parkhust, Law, and Niebur 2002), making it efficient to isolate, there is some evidence that individuals engaged in a visual search task with recognizable, naturalistic images are more likely to be distracted (and therefore have their search response times increased by) similar figure distracters that are visually similar than distracters that are the same color as the target (Williams, Henderson and Zacks 2005). This is despite the fact that individuals can effectively extract color, orientation, and intensity in order to develop an understanding of a visual scene (Parkhurst, Law, & Niebur 2002). This may also have some relation to the Stroop task, where it is easy for individuals to ignore the color of a word, but not the semantic meaning of a word. Within the context of this study, this suggests that visual images that are semantically similar to the target may be more likely to be confused, and increase response time, as opposed to items that are merely the same color (Stroop 1935)

2.6 What this suggests for cueing

The primary role of cues in visual search is to decrease response time when searching for a target (Yantis and Jonides 1984). The generally accepted way that cues accomplish this is by limiting the number of items that are considered. There are several different types of cues that can be used to direct attention to relevant images within a scene. These include salient feature based cues such as color, orientation, motion, or cues to the type of figure (Parkhurst et al. 2002, Hillstrom and Yantis 1994). Spatial cueing has been noted to be particularly effective in decreasing visual search time by a number of studies (see Table 2-1 for a summary of different cue types). Both Soto and Blanco (2003) and Lamy and Tsal (2000) found that individuals preferentially used spatial cues over object features. Shih and Sperling (1996) suggest that the mechanism behind feature-based search is not that individuals successfully eliminate objects of a particular color from the display, but rather that observers use the cues to determine the correct spatial location. For example, a person might locate the color red, then recognize it as a red jacket. Spatial cues decrease response time, with a systematic increase in response time as the target is located further from the cue (Downing and Pinker 1985).

Table 2-1. Difference between types of cues

	Naturalistic Image	Artificial Image	
Spatial cue	Area in space	Area in space	
Figural cue	Learned cue to spatial location	Slower,No learned cue to possible spatial location	Oliva and Torralba (2007), Tatler and Melcher (2007)
Color/features	May be less useful due to complexity of image	May be more useful in simplistic images	Williams et al. (2005)
Contextual Cue	Implicitly known by novice observers	Learned over time	Chun and Nakayama (2000),

Spatial attention to a specific area is necessary but not sufficient for target recognition or change detection. As one might suspect, changes to a static picture are best detected when individuals are looking directly at or nearby the spatial location of the change (O'Regan Deubel Clark and Rensink 2000). However, even direct gaze may not be sufficient for change detection. O'Regan et al. (2000) found that even with direct gaze, observers failed to notice a change in naturalistic stimuli 40% of the time when the change occurred during an eye blink. Others, such as Simons and Levin (1998) found that even if individuals looked in the same area as a scene change, they did not necessarily attend to or notice that target when it conflicted with their goal (see Table 2-2).

Table 2-2. Factors impacting change detection		
spatial attention	Allocation of visual attention to a specific location in space	O'Regan et al. (2000)
Interruption of visual field	Occlusion or temporary removal of stimulus	Simons et al. (2000)
Goal orientation	Objectives and steps to complete objectives	Simons and Levin (1998)
Expertise within a domain	Extended practice with or exposure to a given area	Werner and Thies (2000), Reingold et al. (2001)

2.7 Location awareness

From the perspective of a human observer at eye level, location awareness and experience has been shown to have a dramatic impact on visual search strategies. As mentioned earlier, Brady et al. (2008) found that increased time of exposure to a set of stimuli gradually increased recall for the details of that stimuli. Melcher (2006) found similar results for increasing memory of a visual scene over the course of one minute, even with interruptions of short term memory, indicating details are most likely stored in long term memory. Human location awareness, provided by previous experience observing a specific scene, can therefore increase knowledge of elements within this scene over extended periods of time.

Humans as soft sensors with location awareness preferentially allocate gaze both on the basis of categorical scene recognition and on the basis of specific scene recognition. Torralba et al. (2006) found that observers' visual fixations tended to cluster in similar areas of naturalistic scenes when given instructions to find and count specific types of objects. Observers, for example, preferentially fixated on sidewalks when looking for people in a scene, or at a coffee table instead of a wall when looking for mugs. Observers are also capable of encoding contextually relevant spatial information about a specific scene. Becker and Rasmussen (2008) found that observers who had already performed a change detection task in a given scene preferentially looked in the location where the change had previously occurred, later allocating gaze to the new change (addition of an object) when it was placed in a different location than on the previous trial.

Given the persistence of memory for the likely spatial location of objects, individuals unfamiliar with a given environment are likely to be at a disadvantage, particularly in the first minute of exposure, but possibly for an even longer period of time. The memory for collections of items found by Brady et al. (2008) increased for up to the maximum period of time allowed by the experiment, 18 seconds, and Melcher (2006) found that memory of spatial position of items and item characteristics increased linearly over time, and did not level off for up to 20 seconds. Importantly, Melcher (2006) found that information about a scene persisted even after viewing was interrupted by tasks known to utilize short term memory. This offers evidence that information in naturalistic scenes is in fact retained in long term memory, and is used when accessing information about image characteristics. Location aware devices that help an individual allocate spatial attention directly may therefore help to supplement an observer’s location awareness.

2.8 Questions

This literature review suggests that observers have difficulty searching by feature in naturalistic scenes (color, shape, etc.), and they are likely to use category information and spatial position. This literature review covers some of the difficulties in visual search and how spatial cues may assist in resolving them. There is a relative paucity of experiments considering visual search ability in naturalistic video footage. This thesis considers whether observers' pre-existing ability to find objects given a categorical, verbal description within a familiar visual context is faster than a description with a spatial cue to that object. As with other spatial cues in the literature, the proposed spatial cue would act by effectively limiting the number of items to be considered by directing attention to a specific location.

Chapter 3

Experimental Instrument and Methods

Given the likely relationship between familiarity with surroundings and visual search, Penn State's University Park campus was chosen as the location for the video footage used in this experiment. A portable digital camcorder attached to a tripod was used to capture video footage near the Eisenhower Auditorium on graduation weekend, May of 2008. This served a number of different purposes. The primary goal was to gain footage of a crowd scene, which would contain sufficient numbers of people and objects for visual search to be complex and challenging. Graduation weekend provided a good opportunity to gain footage of people from diverse age groups, as both students, parents, and younger family members were in attendance. While the background remained familiar, the crowd was unusually diverse for a college campus. The time of filming was chosen to maximize the number of people in attendance.

3.1 Footage

The video footage was loaded to a personal computer equipped with Sony Vegas video editing software. The first step in the video clip extraction process was to locate unique items within the crowd that persisted on screen. These items needed to be difficult, but not impossible to find. The difficulty in selecting video clips for visual search tasks is discussed by Dee and Velastin (2008). In a naturalistic environment, all items tend to be occluded at some point during search. In the case of moving crowds, people tended to move through the scene and off camera. It was observed that a sort of crowd horizon emerged. In order to find a unique individual, people had to be able to see a relatively large amount of detail concerning each individual. At the edges of the crowd, many details were visible, but within the crowd, very few identifying marks were visible. Therefore the actual number of people identifiable as targets was constrained to those people who were on the outer edges of a crowd.

Figure 3-1. One of the initial clips tested, target is a man in a tan cap.

3.2. Clip choice and pilot testing

Video clips 20 seconds in length were chosen and pretested for search difficulty (see Figure 3-1 for an example). Individuals in the crowd with uniquely identifiable items of clothing or who were carrying unique items were selected. Initially, a longer time frame would have been preferred, but this was not possible as most people did not stay within range of the recording equipment for an extended period of time. Clips were pre-tested by graduate students in the lab who had not seen the original video footage, in order to narrow down the possible clips using VLC media player. A clip with rapid response time was selected for the practice session.

Pilot testing was conducted using three of the most likely clips, with spatially cued, false cued, and cue absent conditions. Of the clips, the clip containing the man with the camera bag had the fewest observational errors, and the clip containing a man with a tan cap had the most observational errors.

The cued boxes were developed using a mask overlaying the video using Sony Vegas. The eventual cue size consisted of a 0.4 cm red border with a 1.5 x 1.5 interior space. The resulting video footage was rendered to 640 by 480 resolution, which is the standard resolution for analogue television broadcasts. The eventual footage was presented at 28cm wide by 18.5 cm high.

3.3 Participant interface development

The participant interface was developed using Macromedia Flash 8.0. The video files were converted into Flash Video (FLV) files, using the included Macromedia Flash Video Encoder. Actionscript code using the included FLV video controls was used to render the interface responsive to participant input, and to record the amount of time it took for individuals to find the target, and then to click on the target. After participants selected the target, a green circle appeared around their selection, and the location of their selection was recorded (see Figure 3-2). Results from the Macromedia Flash video were sent to an Structured Query Language (SQL) database on a local server. Due to network security settings, it was necessary to run the Flash Document embedded in a Firefox web browser.

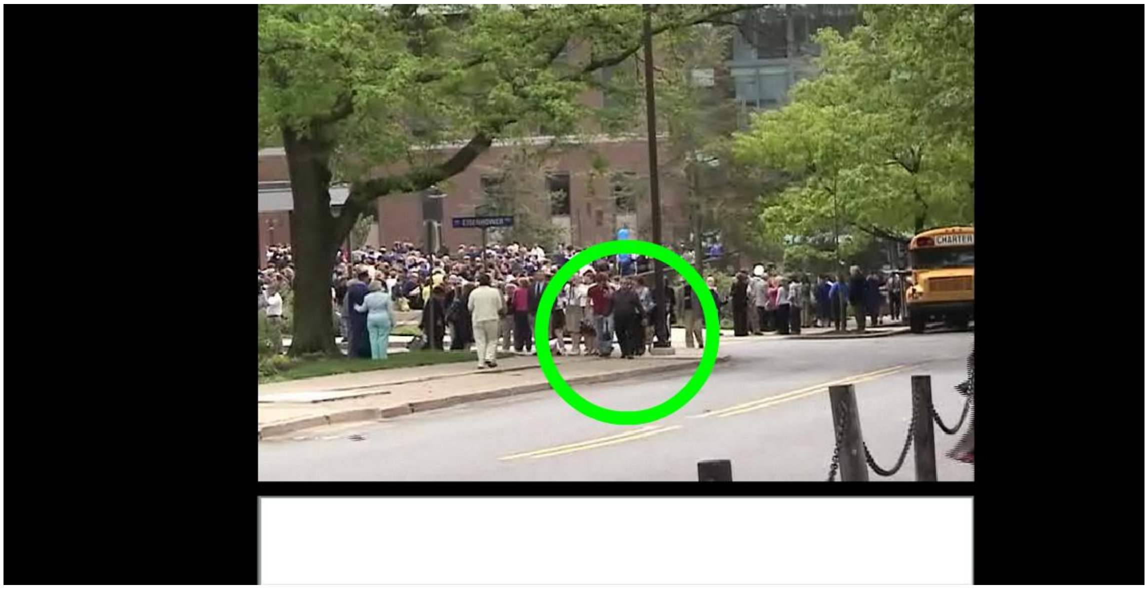


Figure 3-2. Completed visual search task with target correctly identified.

3.4 Survey Materials

Survey materials were constructed to act as a manipulation check for whether participants found the target and to gather feedback for future changes to the attentional allocation tool (red square outline). They were also used to gather general demographics for future analysis regarding gender and age variables, and given the probable relevance of location in visual search tasks, to determine whether or not participants recognized to location of the clip. One question also addresses the issue of whether or not participants recognized any individuals within the crowd, as this was seen as a possible confounding variable.

Method

3.5 General Design

This experiment implemented a four factor mixed design. There were two between subject conditions, cued and uncued, with two random factor conditions, location aware and location unaware.

3.6 Participants

Participants were undergraduate and graduate students at Penn State University with normal or corrected to normal vision. Undergraduate students were offered extra credit through the College of Information Science and Technology classes as well as College of Communications classes. Extra credit was offered through IST 440W and COMM 420. The experimenter verbally recruited participants during class time, then through the email list serve for these classes. The time for participants to attend experimental sessions was arranged through email correspondence. Verbal and email recruiting was also employed in SRA 111, though no extra credit was offered. Email solicitation only was employed through the IST graduate student list-serv and with the permission of the list administrators, through the photography club and the Security and Risk Anaylsis club list-serv at Penn State.

3.7 Procedures

Participants were given a chance to review and sign the consent form shortly after entering the lab (see Appendix A), and were then seated at a computer 40 cm from the screen. No restraints were used, as per previous studies done in the area of computational visual perception (Brady et al. 2008).

The first task was a practice task where participants were verbally instructed to search for a practice target, a woman in a red coat, as quickly as possible while maintaining accuracy. After visually locating the target, they were verbally instructed to press the spacebar, then select the target using the mouse as quickly as possible without sacrificing accuracy. Times were not recorded for this condition, as the experimenter used this session to offer basic feedback on the correct operation of the stimulus (e.g. press the spacebar to stop the video, if target is occluded after pressing the spacebar, click the last known location). Correct responses were verbally verified by the experimenter.

The instructions for the experimental stimulus were the same, except that participants were instructed to “focus their attention on the red box” while conducting the search if there was a red box present on the screen. The target for the second task was to locate a man with a camera bag. After participants completed the search task, the stimulus was minimized and they were given the exit questionnaire in paper format with their randomly designated participant number attached to the corner of the questionnaire materials (see Appendix B). In the case that participants incorrectly located the target, or did not locate the target, they were informed after completing the task, but before completing the exit questionnaire so that they could consider factors that may have led to the incorrect choice on the exit questionnaire.

3.8 Independent Variables, Dependent Variables, and control Variables

The independent variables in this experiment were the cued vs. non-cued conditions and the non-cued expert condition. The dependent variables were time to stop the video, and whether or not the participant recognized the location of the crowd.

3.9 Hypothesis

H1: The cued condition will have a faster response time than the cued condition.

H2: The response times for the cued condition will cluster around the time at which the target enters and leaves the cued location.

H3: Observers who recognize the location of the crowd will have faster response times.

3.10 Data Analysis

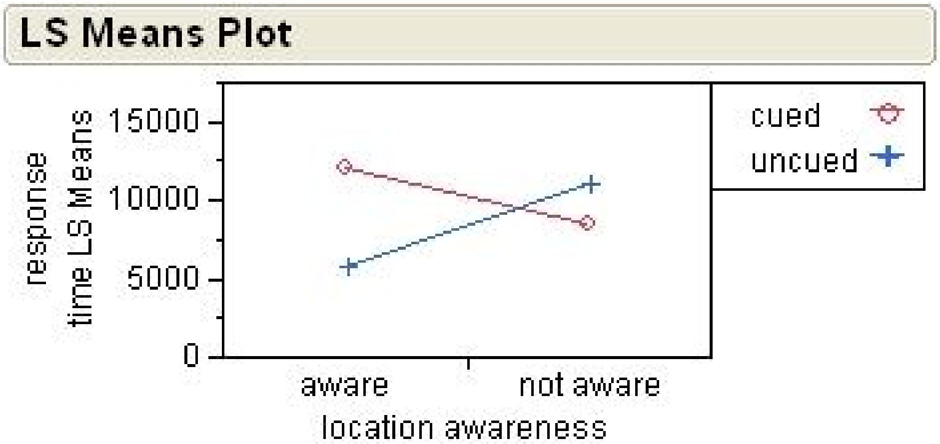
The results of this experiment were analyzed using an ANOVA in order to determine the difference between the cued, uncued, and location aware conditions. Participants who failed to see the target or incorrectly located the target were excluded from an analysis of response times.

Chapter 4

Results

4.1 Quantitative results

Out of 34 participants, 7 participants failed to correctly locate the target. Of those participants, 5 were in the cued condition and 2 were in the uncued condition. Of the people who failed to find the target, 2 out of 7 recognized the location of the crowd. The target was presented in the cue box between 10 and 11 seconds into the clip, and 5 of the response times were within one second of the cued time and place (see Figure 4-2). A 2x2 ANOVA (cued x location awareness) showed that the effect of cueing (uncued M = 9, SD = 5064; cued M = 9796), SD = 3645) and location awareness (aware M = 8879, SD = 5566; not aware M = 9331, SD = 3076) had no main effects ($F(1, 19) = 1.29$, $p = 0.27$; $F(1, 19) = 0.32$, $p = 0.58$). However, there was a significant interaction effect between the cueing and location awareness $F(1, 19) = 7.01$, $p = 0.016$. While the effect is significant, this data should be considered preliminary pending further participant recruitment, as the number of participants per cell is slightly below the amount generally recommended for parametric statistical testing.



Means for awareness and cueing (in ms)		
	Aware	Not aware
Cued	12025	8558
Uncued	5734	11070

Figure 4-1. Graph and table showing interaction of location awareness and cueing.

For participants who were aware of the location, the cue negatively impacted performance as measured by response times, but for participants who were not aware of the location, the cue improved performance as measured by response times (Figure 4-1).

Excluded data

Out of 34 participants to complete the experiment, 11 were excluded from the analysis of response times. 7 of those excluded failed to complete the task, and missing data was present for the location awareness of 4 participants.

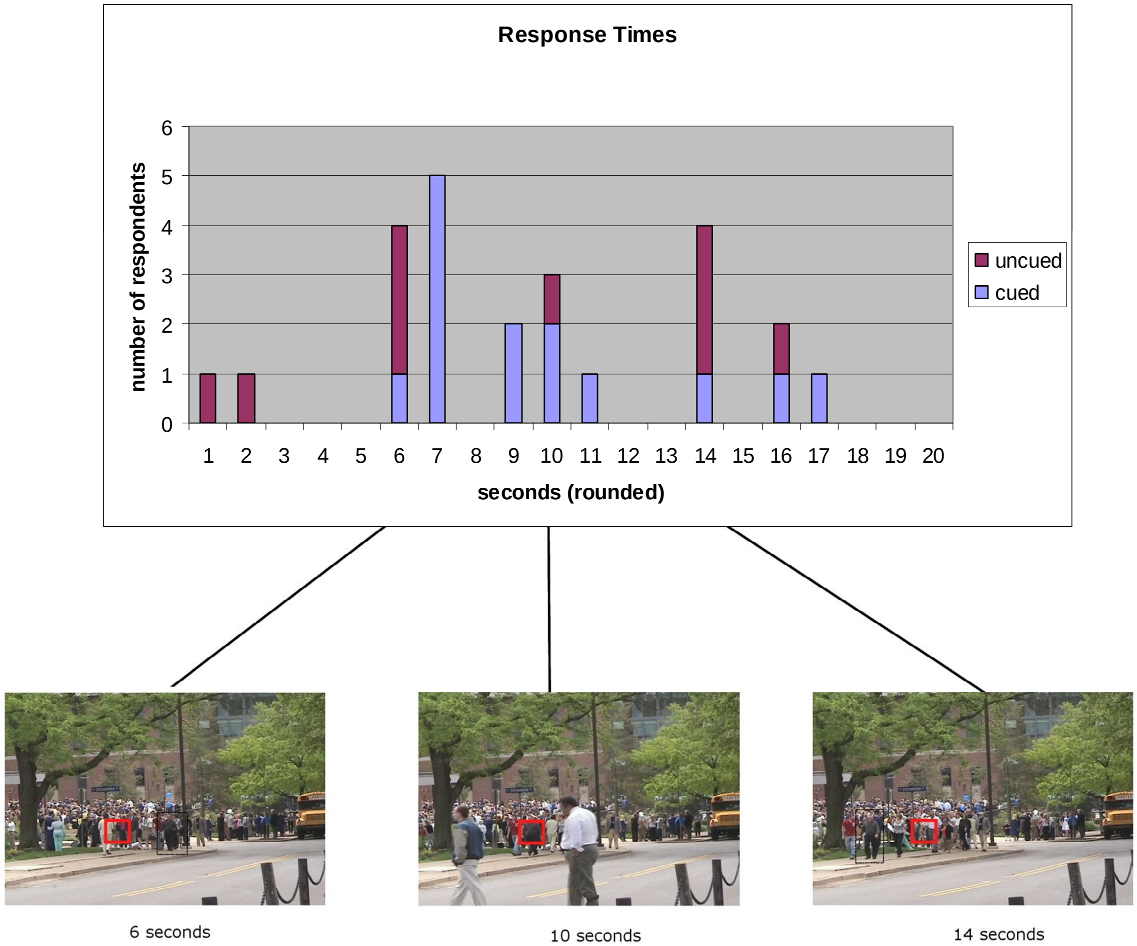


Figure 4-2. Spatial position of target at different times.

4.2 Questionnaire results

The questionnaire was designed as a manipulation check, to collect data on aspects of the stimulus that may have led to participant error and incorrect response, and to monitor for potentially confounding items. It also served to collect the random factor, location awareness.

The first item was a record of whether or not the participant correctly located the target (the experimenter checked responses after data collection to verify that they had correctly filled out this section). One participant responded that she located the wrong bag, and another responded that she “just clicked on the red square because I assumed it was there". This participant was excluded from further analysis due to failure to understand directions. The other respondents replied with yes/no answers.

The second question, whether or not participants knew the location of the crowd, tested the random factor, location awareness (10 aware, 13 unaware included in the final analysis, see Figure 4-3). The majority of the responses were yes/no responses. Any response that indicated familiarity with the location was coded as knowing the location of the crowd. Out of all the participants, 5 responded by identifying the location using the term "Eisenhower", one participant responded by stating "Penn State's campus" while the rest gave yes or no answers. There were no responses incorrectly identifying the location as somewhere other than the Penn State Campus.

The third item was intended to control for the potentially distracting effect of recognizing a person in the crowd. No participants reported recognizing any individuals in the crowd prior to the experiment.

The fourth item, measuring whether the participant found the attention directing tool to be distracting, was a manipulation check item to see if individuals found their attention effectively directed by the tool as well as an item intended to provide information regarding the use and attributes of the tool.

The survey data generally indicates that the attention cueing tool effectively drew attention to the area of the screen it represented. One participant noted "the corners drew my eyes like an arrow would", while another suggested "I found myself looking there instead of scanning the crowd". The most common complaint in the surveys was occlusion. One participant noted that the square partially obscured the camera bag, while another noted that the car obscured view during the course of the clip, which was accurate. There were only two false positive responses during testing; participants tended to either miss the target entirely or accurately pick it. One participant incorrectly chose a woman's black purse instead of the target (cued condition), while another participant incorrectly chose a tan purse instead of the camera bag (uncued condition).

In general, participants who failed to find the target did not elaborate on why they failed to find it. Of the participants who failed to recognize the target, one noted that "The red box was distracting along with the fact that people were walking in front of the person in the red box" showing participants' awareness of occlusion.

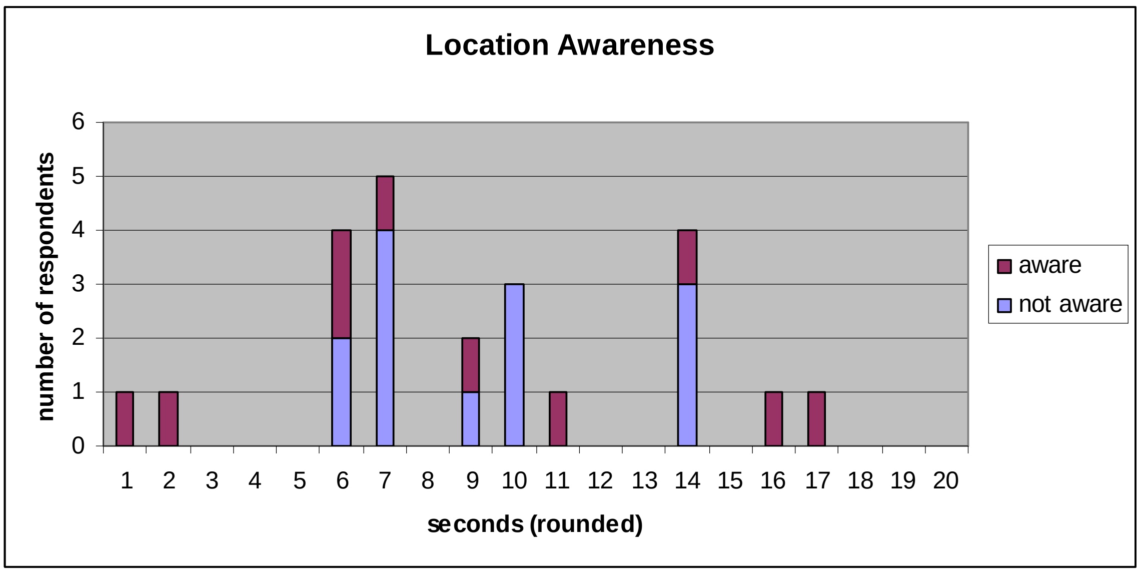


Figure 4-3. Graph of location aware vs. location unaware participant response times.

Chapter 5

Discussion

5.1 General Discussion

The primary question addressed by this experiment is whether spatial cueing effectively reduces the amount of time required by novice observers to find a target stimulus in a motion picture, and whether it increases the likelihood that they will find it. The evidence suggests that if participants recognized the location, they performed worse on the cued condition for the clip that was tested. In contrast, the spatial cue was useful to participants who had no previous reported experience with the scene. This supports the literature on contextual cueing in naturalistic situations (Oliva and Torralba 2007, Torralba et al. 2006, Becker and Rasmussen 2008).

There are several factors related to the individual video clip that may have influenced participant performance. These include the time at which the target was cued, sudden occlusion by moving elements in the video clip, and the placement of the cue relative to the attended elements of the target. Part of the observed effect may be due to the spatial allocation cue directing gaze to the correct location later in the clip. The target was presented in the cue box between 10 and 11 seconds into the clip, and 5 of the response times were within one second of the cued time and place.

The probability of finding the target was not equal at all times. The target was at times occluded or difficult to distinguish from a background of other individuals in the crowd (see Figure 4-2). While there was not a significant difference in the means for the cued and uncued condition, indicating that the response times for two groups did not differ significantly based on the cued location, however, the data show spikes in uncued observers correctly locating targets at 6 and 14 seconds, as opposed to 10 seconds, which designates the time the target passed through the cued location (see Figure 4-2, 5-1).

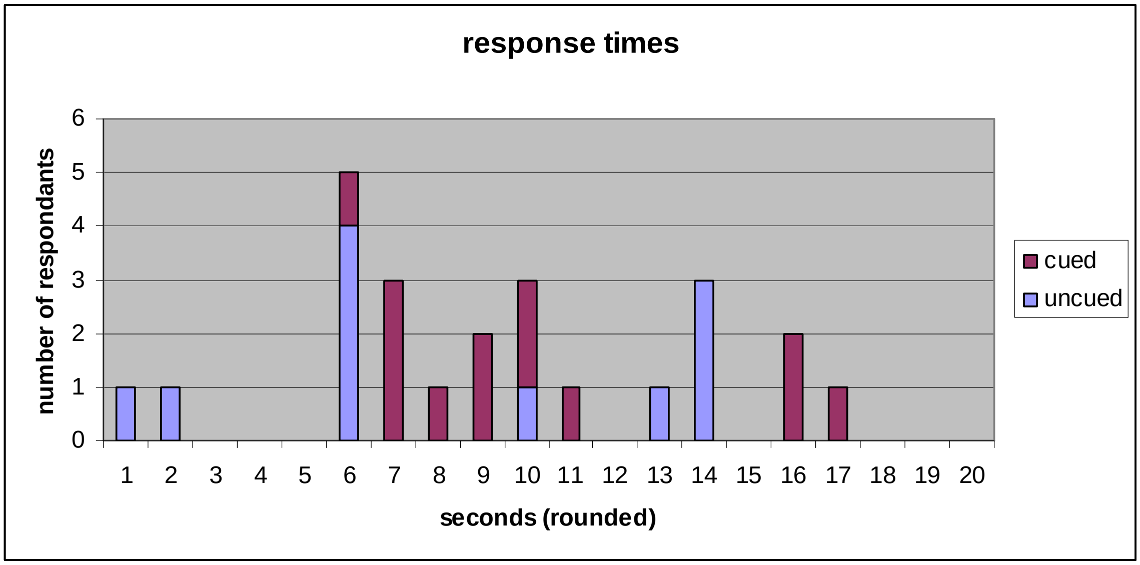


Figure 5-1. Response times for cued and uncued observers

The manipulation appeared to be successful. A manipulation check was included in the questionnaire in that participants in survey data indicated that the attentional allocation tool effectively directed their attention to the given spatial location. Observers in the cued condition were more likely to locate the target between 9 and 11 seconds than those in the uncued condition, with only one uncued observer locating the target during this time span, and 5 cued observers locating the target during this time, although this is not a reliable difference (Figure 5-1). This suggests that the target did in fact cue observers to the target during this time span.

While the tool effectively increased target detection for the time that it cued the target, by not cueing at the locations where the target was most visible, the effect was contingent on observer awareness. Cues of this type are known to effectively redirect visual gaze in naturalistic video (Dorr 2004). Observers who were aware of the location had their attention effectively drawn to a location that was not as effective as one that they might have picked based on their own experience. Furthermore, the attentional direction of the tool may in fact have been overriding, directing observers not only to the location it pointed towards, but to the cue itself given the salient features of the tool (Thueewes 2004).

While the literature suggests that spatial cueing is highly effective, as it appears to have been in this experiment, there is a cost in terms of attentional breadth. Spatial cues draw attention away from the target in other spatial locations, potentially acting as distracters when they are inaccurate, unclear, or spatially inappropriate to the target. In previous research, cueing of one target has been shown to detract from the ability to find later targets, or unexpected targets (Yeh and Wickens 1998). This represents a contradiction; in order to effectively cue an observer to a location, the cue must be salient and easily found. In some studies, a spatial cue is shown briefly before the stimulus is presented, eliminating this problem. While it is not difficult to present a cue to spatial location immediately preceding a static image, surveillance footage is a continuous phenomena. It would require interruption of the footage to do so, a known risk factor for increasing change blindness (Simons et al. 2000). Dorr (2004) used a rapidly appearing and fading visual stimulus in order to elicit gaze to a specific area of naturalistic video, which is another possible option.

Some of the feedback received in the context of this study suggested that the man carrying a camera bag was difficult to resolve due to partial occlusion while in the cued location (see Figure 4-2). Spatial cueing is less than effective if the visual image is difficult to resolve. In the case of thermal imaging sights, soldiers have failed to effectively use weapons despite thermal guidance systems using a spatial cue, a sight, due to an inability to identify the images within the sight (Blackwood et al. 1997). Spatial cueing only points toward a given location, it does not aid in identification per se.

Occlusion and resolution of the target may not be the only issue facing observers viewing a moving video. As evidenced by the literature on “mudsplashes”, the sudden addition of salient elements to a display can distract observers from other changes in a display that occur at the same time (O’Reagan et al. 1999). Given that the target was clearly visible for at least part of the time he was in the cued location, and multiple participants found the target between 7 and 11 seconds, it seems unlikely simple occlusion completely explains participants’ difficulty in finding the target in the cued condition.

A more effective interpretation might be that other objects moving across the visual field acted as distracters, and that visual attention to the area of the screen where they appeared increased the distraction effect. There is a time period after allocating attention to a suddenly appearing object and identifying it during which an observer is unable to redirect attention and identify subsequent items. This effect, referred to as the attentional blink, lasts on average, between 200 and 500 ms (Raymond et al. 1992). If participants directed their gaze to the cue box for the entire course of the experiment, this effectively decreased the time period they would have been able to find the target, and magnified the effect of the attentional blink caused by objects rapidly entering their field of view.

The cue used in this experiment is similar to a crosshair or sight used in a video game. One compensatory mechanism to render salient stimuli in foveal vision less distracting may be to render them stationary in the visual field, decreasing the motion salience. Eye tracking data from video game research indicates that players in first person shooter games tend to move the crosshairs and focus of the game rather than move their gaze relative to the monitor even when examining non-target items (Holmberg 2007, El-Nasr and Su 2006). One explanation for this may be focusing the center of vision on the center of the screen effectively immobilizes the non-changing parts of the display such as the health meter and crosshair in the visual field, making them less salient.

Notably, this is less true in games that do not require fast responses, and that do not have a crosshair (El-Nasr and Su 2006). El-Nasr and Su (2006) also found that salient changes to the environment such as luminance and color attracted visual gaze, but not when it would override real-world based strategies such as focusing on doors to find an exit. In this case, the changes were only noticed when they occurred spatially nearby a door. This task is distinct from a crowd search task, as the certainty of finding the correct target in different parts of the crowd is less than that of using a door as an exit, but it does show that in some cases, goal directed visual search is likely to override a salient spatial cue. This suggests that contextual cues relevant to a goal are likely to overwhelm salient cues, in this case, the red box action as a spatial cue.

One of the questions implicit in this thesis is how observers who correctly located the target identified him, and why those who did not failed. Overall, 5 participants failed to find the target, and 2 incorrectly chose a target. This indicates an overall 79% accuracy (Figure 5-2). The clip lasted 20 seconds. Research on change blindness indicates that approximately 3 seconds per item is necessary in order to build a detailed representation of an item. While more than six items appeared in this video, there are a number of reasons that participants were nonetheless generally accurate in the search task. The first was the salience of the target: The man with the camera bag was walking through the scene. While there were multiple moving objects in the scene, his motion would have effectively separated him from the static background (Hillstrom and Yantis 1994).

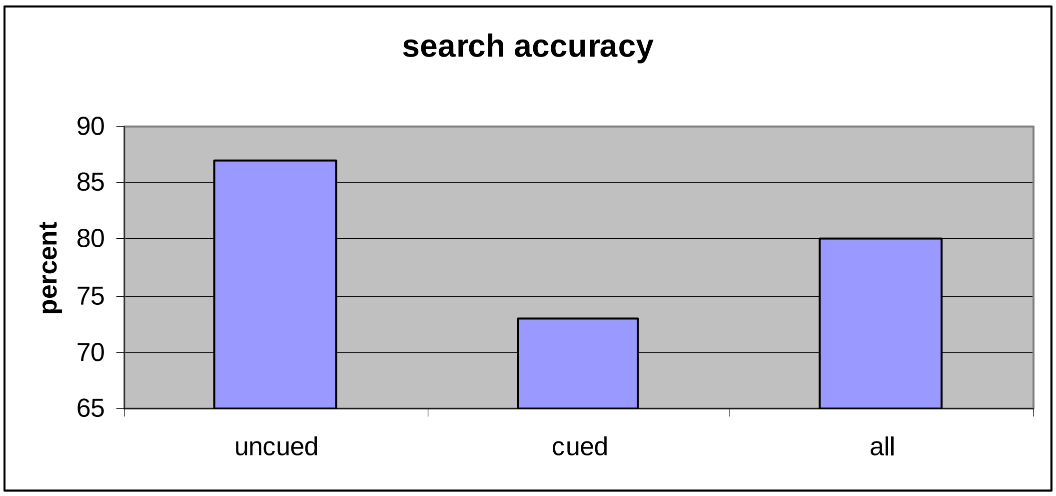


Figure 5-2. Percent of participants who successfully completed the task.

It also may not have been necessary to fully represent the target in order to successfully locate him. The results of Brady et al. (2008) suggest that while it takes seconds to encode complex information about an object (such as whether a mailbox flag is in the up or down position), categorical information recognition and information retention requires less than a second (mailbox vs. rubber duck). This agrees with the results of Rousselet et al. (2005) who also found very rapid categorical recognition for naturalistic images (landscapes). As predicted by this theory, all errors corresponded to items of a similar category, in this case, bags.

Extensive practice is not required to improve novice search and recognition of specific categories of items. In order to improve target recognition, a simple training session containing items of a category similar to the target may allow the human as soft sensor to more accurately find and discriminate between targets. GalaxyZoo is one example of a website which tasks novice observers to find and categorize visual targets (Lintott et al. 2008). In this example, Lintott et al. (2008) solve the problem of contradictory findings by humans as soft sensors (Pravia et al. 2008) by taking a weighted sample of novice-identified galaxies. Classification for direction of spin agreed with expert observers on over 99% of cases, with novice users required only to identify 11 out of 15 galaxies correctly on a quiz before beginning to classify (Lintott et al. 2008).

Experience with location as well as the target category may have potential to increase the efficiency of visual search. The best observers were those in the uncued, location aware category. While the spatial cue improved performance for those who were unaware of the location, it did not improve the response time of location aware observers. Observers who were able to recognize the location of the crowd generally performed slightly better than those who did not (in terms of correct vs. incorrect respondents), and all five of the individuals who recognized the location of the crowd as Eisenhower specifically correctly located the target. This may indicate that by knowing the location of the clip observers were better able to employ more effective strategies, or it may simply indicate that these individuals were better observers in general, and their ability to quickly recognize video of a given area is tied to their ability to find a target.

Recognizing a type of location has been found to bias gaze toward certain spatial locations within that area (Chun and Nakayama 2000). In this case, recognizing the location of the video moderated the effect of spatial cueing on the response time. Location awareness may have allowed observers to more effectively allocate spatial attention, enabling them to find the target before it would have entered the cued location. In Figure 5-3, response times for observers who recognized the location are more broadly spread out than those for observers who failed to recognize the location of the crowd.

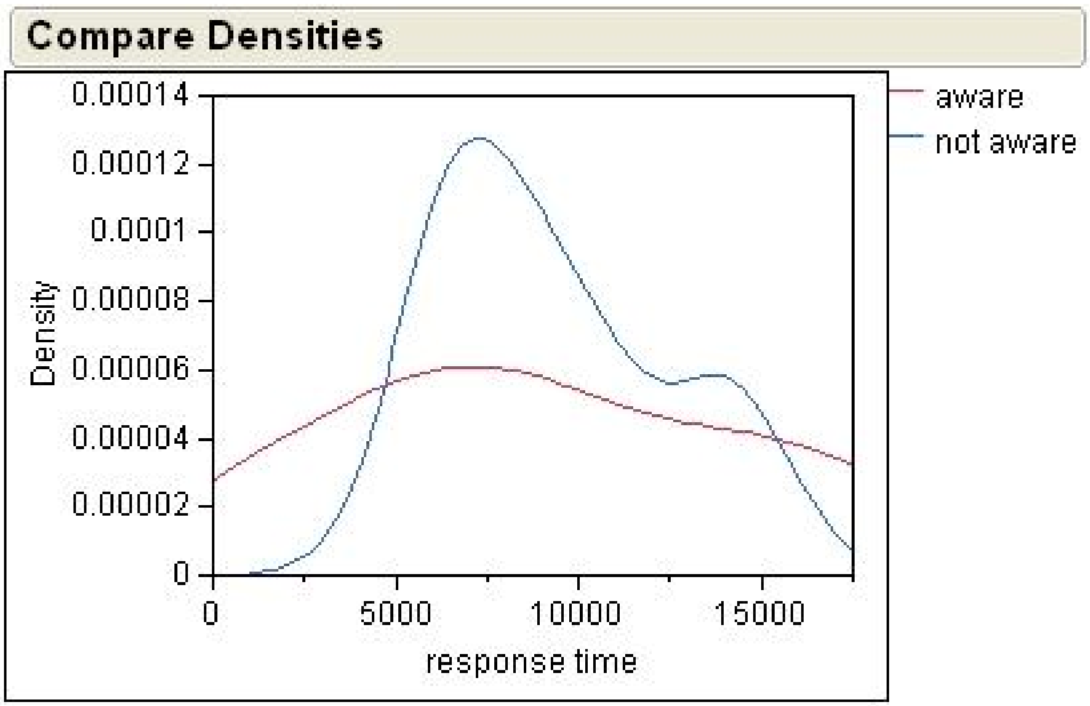


Figure 5-3. Graph of densities of response times for location aware vs. location unaware conditions. Response times for the aware condition are distributed more evenly over time.

The response times for the location aware and location unaware condition were not significantly different, but a graph of the response time densities seems to suggest that the response times clustered differently for the two groups (Figure 6-3). Response time indicates both a temporal and spatial location, therefore response times that indicate a central time also indicate a more central location as the target walked across and then off the edge of the screen. Observers who recognized the location may have been more willing to attend to different locations within the scene while those without direct knowledge of the area's spatial layout may have adapted a different strategy, searching more in the center of the scene.

Despite the fact that the location contained noticeable landmarks, and all participants were students at the University Park campus of Penn State University, many participants were unable to cite the location of the video clip. This is an expected result congruent with the existing literature on inattentional blindness. It should be noted that University Park is a large campus with over 200 buildings, though filming took place near several well-used academic buildings, an auditorium, and down the street from the Creamery, a popular campus location. It is therefore possible that some students had not viewed that particular area of campus, or not viewed it from that angle.

5.2 Limitations

The primary limitations of this study were the small number of participants, the use of a single video clip, and limited feedback from questionnaire items. This study also employed a between subjects design in order to minimize practice effects. It may be that a stronger effect for the cued vs. uncued conditions would emerge in a within subject design. Given the interaction effect showing a relationship between observer differences (location awareness), a within subject effect showing a decrease in performance for location aware observers in a cued condition compared with an uncued condition would be expected to emerge. Given a larger number of participants, main effects for the location aware and unaware condition may also have emerged.

Only one video clip was used in this experiment due to participant limitations. In order to properly counterbalance for order of presentation, 60 participants would have been needed for only 2 video clips. Of the clips initially tested, the clip with the fewest failed responses was taken to minimize the number of participants to be discarded. Other types of data analysis, notably ordinal regression, would have allowed discarded clips to be directly compared to successfully completed response time, but at a loss of data granularity (Hedeker and Gibbons 1994). A collection of video clips with pre-tested, average response times would be a highly useful resource in testing the effects of different cueing devices. In this experiment, the level of difficulty and individual characteristics of the scene and target likely impacted the eventual results.

One of the cued condition clips used in pilot testing might be more effective at drawing user gaze to the target with user identification of the target than the clip used in this experiment (Figure 3-1). There are a number of possible reasons for this. The cue box may be at a more spatially appropriate scale to capture the visual element being used for discrimination, (Nothdurft 2002), and it captures the target in a non-occluded state. The difficulty of the clip may also be greater, which would lead to a higher number of individuals failing to find the target entirely, but would highlight the utility of the cue.

For future research, the questionnaire instrument should more clearly specify that participants should list the name of nearby buildings or street names as location markers, and ask participants if they were aware of the specific intersection. Responses indicate that it may have not been clear what type of location information the questionnaire required, and this limits the construct validity of location awareness.

5.3 Future Research

For future research, the effects of several factors should be considered. The characteristics of the cues used in visual search research vary (Nothdurft 2002, Sadasavian et al. 2005, Dorr 2004). Different characteristics of these cues, including their shape, temporality, and relationship to search strategies would be interesting grounds for further investigation. Previous research suggests that individual differences such as video game experience may also have an effect on visual search ability. Security experience has not been shown to have an effect on the ability to recognize developing problematic situations in video (Troscianko et al. 2004), however, this task is more specific in its focus, and perhaps more analogous to finding a lost individual than detecting a non-specific threat.

One of the benefits of choosing a between subject design is that future research utilizing different independent variables but similar methodology can compared to the current research. The current data can be used to determine average response times and compared against future conditions. One of the conditions that might be modified in future research is the characteristics of the cue. The cue itself should be modified to determine those effects. The red box cue was found to be occlusive by at least one participant. To avoid any problems with cue obfuscation, for future research a thinner, partially transparent cue might be more effective. There may also be effects for a cue surrounding potential targets as opposed to a solid marker, as a number of participants commented that they tended to confine their vision to the area within the box, although they were not specifically requested to do so.

There is research suggesting both increasing and decreasing the size of a surrounding cue may have an effect on reaction time. Previous research by Nothdurft (2002) has employed a simple shape surrounding a target, which was more effective when the circle closely surrounded the target. Nothdurft (2002) suggests this is due to both being on the same spatial scale, but another explanation might be that it simply drew attention closer to the target, decreasing the number of items necessary to search. In this experiment, the cue surrounded the target and bag, but it appears from participants who made errors on the visual search task that the camera bag may have been the primary way that the target was identified, with the person carrying it receiving little attention.

This is in contrast to other literature which suggests that the larger the area of attentional focus, the more accurate the change recognition is. Both research on effective field of view and expertise suggest that increasing the area that observers are able to continuously monitor increases the likelihood that they will detect change (Caird et al. 2005, Reingold et al. 2001). It may be worthwhile to design and test a cue to allocate attention to encompass the maximum area that an observer is able to view. Part of the reason that the search task may have been relatively easy is that is did not extend over a large field of view. For larger visual scenes, and more difficult tasks, it may be more effective.

Spatial cueing may be more effective if it is illustrative of a goal based search pattern rather than a single point on the screen. Previous studies have indicated that studying representations of search patterns based on eye tracking data may be more effective for aiding visual search than single highlighted elements within a scene. Sadasavian et al. (2005) note that novice observers found more targets after studying the visual search patterns of experts than after studying only the highlighted areas of importance. Eye tracking data represents a valuable tool that could be used as a more complex spatial cue.

The literature suggests that observers who are expecting a target in a given spatial location are more capable of ignoring other distracters (Yantis and Yonides 1990). The results for individuals in the cued condition seem to suggest that pre-testing for any cuing mechanism in a visual guidance tool should be conducted to measure potential distraction effects. It may be that in naturalistic situations, spatial cues are obtrusive for relatively easy visual search tasks.

Given the apparent importance of context and experience in visual search, for future research, I hope to recruit police observers or other security personnel with expertise in isolating individuals of importance from a crowd. Previous research has generally indicated that expertise in visual search is constrained to a specific domain. Football experts, for example, are more likely to notice changes in configurations on the field but not in the stands (Werner and Thies 2000). The experience of individuals in security or first responder professions may be unusually generalizable given the similarity of their domain expertise to commonly occurring situations, allowing them to develop visual search techniques tailored for many different environments. Previous research suggests that experienced security personnel do not perform more effectively than experts when searching for a non-specific threat (Troscianko 2003), but this study did not give participants specific target features to search for.

Video game experience is another factor that is likely to significantly impact visual search performance. Video game experience has been shown to increase discrimination of targets from distracters both in the center of vision and in the periphery, as well as increasing the useful field of view, a factor known to reduce change blindness (Green and Bavelier 2006, Caird 2005). It may also increase the ability to discriminate targets specifically from crowded backgrounds (Green and Bavelier 2007). The genre of first person shooter games and action video games have shown the potential to increase expertise in ways that is generalizable across tasks. After only ten days of action video game practice for sixty minutes a day, male and female participants increased their ability to count briefly presented objects, their ability to notice new changes to a display after a previous change, decreased the time span of the attentional blink, and increased their useful/effective field of view. Video game experience should therefore be a potential control variable in future research on visual search performance (Green and Bavelier 2007).

5.4 Contributions

Response time itself is an important factor in Network Centric operations, as it aids in establishing a current common operational picture (Wesensten et al. 2005). One of the aims of the project was to measure situational awareness at level 2. At least one tool discussed by Wesensten et al. (2005) has been developed specifically to assess the ability of human as sensors based on simple visual response time, and this work theoretically expands on that work to establish a tool with greater ecological validity.

Research on spatial cueing in general is likely to be relevant to multiple domains, including various biomedical applications, user interface design and visualization of information (see Table 5-1). Spatial cueing in naturalistic video is likely to be specifically relevant to situations which involve rapidly changing visual contexts such as aircraft operation and motor vehicle operation. The current research generally indicates that spatial cueing should be used with caution in environments where a moving target may be present nearby rapidly changing distracters, and that the details used by observers to distinguish the target should be carefully noted so as to effectively cue the details used by the observer

Table 5-1. Domains of Relevance		
Police/Security	Surveillance monitoring	Pravia et al. (2008)
Operation of motor vehicles	Relevant to Effective Field of View (EFOV), Change blindness at intersections	(Caird et al. 2005)
Medical Applications	Radiology (finding areas of low contrast), Sonography, laparoscopy	Myles-Worsley, Johnston, and Simons (1988), Bavelier et al. 2008)
User interface design	Display utilization can be conceptualized as a type of visual search	John and Kieras (1996)
Aircraft operations	Identification of ground based objects, identification of change in visual field	(Wickens and Yeh 1998)
Visualization of Information	Finding information in a large visualized data set	Healy and Enns (1999)

5.5 Conclusion

The practical recommendations for better use of humans as soft sensors in this study focus on the importance of location-based knowledge and training on possible attention allocation tools. For individuals with location awareness, verbal cues may be sufficient to enable effective search strategies. One participant commented that he thought the verbal instruction to look for a camera bag was vague, and that people might often mistake the camera bag for another item entirely, although he was able to locate the object himself. Out of 34 participants, 27 were able to locate the suggested item within the time frame. In general, novice observers given verbal direction or verbal direction with a cue were accurate more than 50% of the time, with only 2 false positives.

Humans as soft sensors need not be people with extensive training in the reporting of information. Expertise within a given domain appears to be more important than experience in reporting information per se. Part of the aim of recruiting humans as soft sensors is to utilize the increase in communication technology to gather information over a broader range geographically and also from a broader range of people. While this increases the problem of filtering messages, it also potentially increases access to individuals with expertise. The literature in change blindness suggests that part of the problem with noticing significant changes or targets in a scene is familiarity with the normal state of those targets (Brady 2008).

Some of the practical implications of this work include tasking humans as soft sensors who have pre-existing knowledge of a given environment or type of environment to report changes. One proposal has been to task sanitation workers (Glater 2009). Sanitation workers have the advantage of long term knowledge of a given environment. They are likely to encode state as well as categorical information that they encounter. For example, they may notice that the car in front of a house seems to change every week. Thus far, sanitation workers have been granted awards for effective services, including awards for noticing suspicious content in a car on a route, assistance in several house fires, pursuing individuals suspected of illegal dumping, and reporting drug related activities (Turso and Duckett 2004).

While tasking underutilized resources such as sanitation workers may represent a valuable opportunity, this study offers a number of caveats when working with untrained observers. While the attentional allocation tool may offer some advantage to individuals unfamiliar to an environment, it also may increase the effect of distracters. Given the apparent distraction effect of the spatial cues in this instance, it may be worth investigating whether these cues would become less distracting with constant use. There are a number of visual errata that observers are constantly faced with in their everyday environments, and these vary from the edges of a new pair of eyeglasses to the constant presence of crosshairs in a video game.

As with other research in this area (Dorr 2004), one of the motivations behind this experiment was to task observers to a specific location. Given the premise that spatial attention is necessary but not sufficient for finding a cue based on the existing literature, it was necessary to test the ability of a cue not only to allocate attention to a given area, but also to test the ability of that cue to aid an observer in identifying a given target. As expected and shown by previous research, a highly salient cue does attract attention (Dorr 2004), but does not always improve visual search in naturalistic video.

New variables that may have a significant impact on the effect of spatial cueing in naturalistic environments were isolated as a result of this study. These include the relative visibility of the target at different points in a video clip and the cueing of specific target elements that participants used to make judgments. In this case, participants may have been searching particularly for the camera bag, and therefore the camera bag itself should have been cued in a spatially appropriate manner. Based on questionnaire results and clustering of response times, it appears likely that the spatial cue worked to allocate visual attention in this experiment. While the hypothesis that the spatial cueing mechanism would lead to grater accuracy and better response times was not shown in this particular experiment, the underlying premise that spatial attention is necessary but not sufficient to complete a visual search task was proven to be accurate.

One of the stated goals of combining soft sensors with hard data in Network Centric Cognition is finding and tracking a moving target (Pravia 2008). In this experiment, the ability of humans as observers to locate a target is tested with categorical verbal cues alone or in combination with spatial cues. It should be noted that even with categorical verbal cues, and no prior formal training concerning those cues, humans as soft sensors in this experiment were generally accurate. Of those who were not, more declined to guess than guess incorrectly. As with similar experiments where novice observers are asked to perform a search and identification task (Galaxyzoo 2009), the majority opinion would have yielded accurate results. Relatively few (6 out of 34) of the observers placed the location of the video footage, however, no participants who guessed as to a specific location were incorrect. While this suggests that items in future research should be more specific in their elicitation of information, it also suggests that in the absence of direct elicitation of location information, humans are observers are unlikely to guess a location without a reasonable degree of certainty.

The general success of participants in an environment familiar to them argues for the use of students on a college campus as observers. In a small crowd, participants given a categorical verbal cue are likely to focus on the same object. Throughout pilot testing of video clips and in the final experiment, false positives were relatively rare. Lossau and Stotz (2008) suggest that the fusion between hard and soft sensors is a process of exchange. In this case, humans acting as soft sensors by observing a crowd situation receive information regarding an important type of visual phenomena, then use situational knowledge to more accurately track a target.

In this experiment, observers pinpointed an exact location on a computer screen corresponding to coordinates on the video frame. This information could then be used to redirect hard sensors to examine the potential target as discussed by Pravia (2008). The most analogous situation would be to task college students to find targets through webcam footage. However, humans as soft sensors could also be leveraged by tasking observers to create a sensor network which would directly transform tracking information to hard data (Hall and Aungst 2008). A sensor network in this case would consist of humans tasked as soft sensors relaying visual information that they collect. In a naturalistic situation, observers would relay video or still images from their current location after locating what they perceived to be the target. Rather than analyzing pre-recorded information as was the case in this experiment, humans as observers would directly act to create new hard data after using their situational knowledge to establish an appropriate location (Figure 5-4).

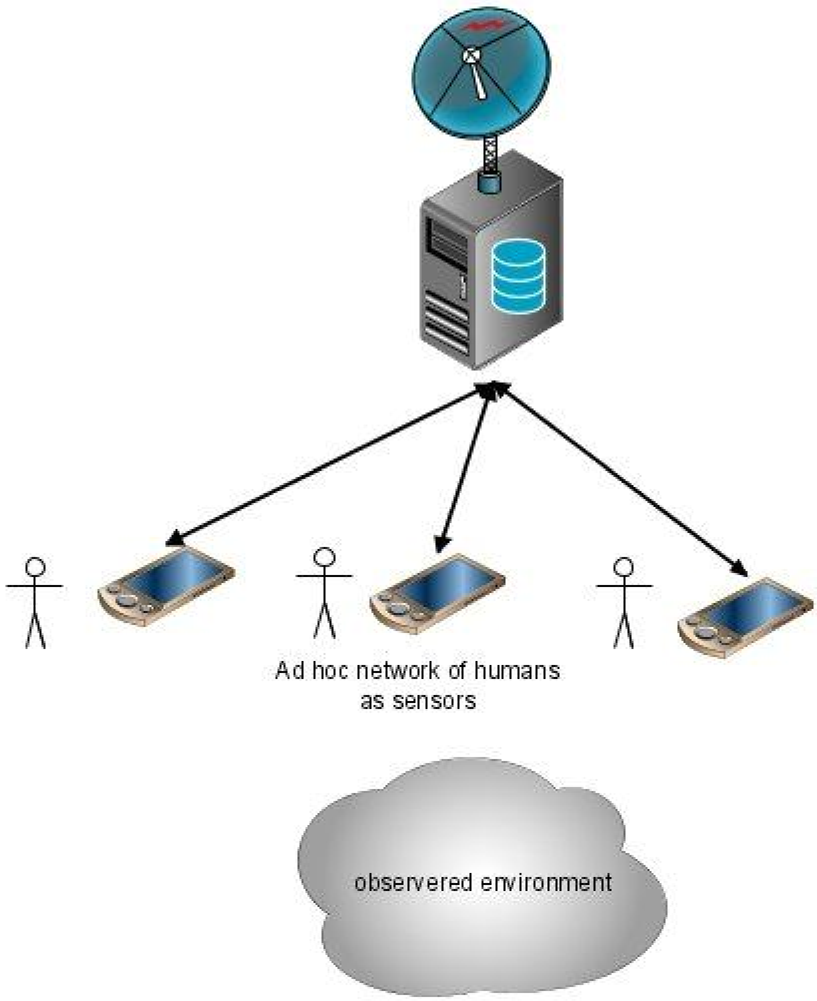


Figure 5-4. Humans acting as soft sensors in an ad hoc network.

Humans acting as an ad hoc network of soft sensors to report information on their own environment are a phenomenon which has already occurred in several instances. In crisis situations, text based information spread through informal networks of individuals in online environments appears to be surprisingly accurate. After the Virginia Tech shootings of April 16, 2007, students and other individuals correctly compiled lists of victims online using Facebook and Wikipedia (Palen et al. 2007). Within seven hours of the first shooting, various lists online, when compiled, revealed a correct list of all 32 victims, with no false positives (Palen et al. 2007). The official list was not publicly released for several hours after this point. Clearly, some individuals in the network had access to correct information, and the informal social network composed of various Facebook groups, Wikipedia entries, and peer-to-peer information spread through cell phone use and instant messaging allowed a rapid conglomeration of information (Palen et al. 2007). During the event, video footage of the actual shooting was taken by a student with a camera phone who was nearby at the time, which also became widely available (CNN 2007).

The results of this experiment generally indicate that an untrained group of individuals with experience in the environment that they observe could provide accurate information when given a specific task. Human observation requires special considerations due to potential sources of distraction, blindness to change, and individual differences in the ability to interpret and understand events due to differences in experience. Moving forward, the presentation and organization of information through technological means is likely to have a significant impact on the utility of humans as soft sensors.

References

Achtman, R. L., Green, C. S., & Bavelier, D. (2008). Video games as a tool to train visual skills. *Restorative Neurology and Neuroscience, 26*(4), 435-446.

Aivar, M. P., Hayhoe, M. M., Chizk, C. L., & Mruczek, R. E. B. (2005). Spatial memory and saccadic targeting in a natural task. *Journal of Vision, 5*(3), 177-193.

Antony, R. T., & Karakowski, J. A. (2008). *First-principle approach to functionally decomposing the JDL fusion model: Emphasis on soft target data.* Paper presented at the Information Fusion, 2008 11th International Conference on.

Apple. (2009). iPhoto '09: Improving face recognition results. Retrieved June 23, 2009, from http://support.apple.com/kb/HT3442

Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception & psychophysics, 55*(5), 485-496.

Bacon, W. F., & Egeth, H. E. (1997). Goal-Directed Guidance of Attention: Evidence From Conjunctive Visual Search. *Journal of Experimental Psychology Human Perception and Performance, 23*, 948-961.

Bai, M., & Ullman, S. (1996). Spatial context in recognition. *Perception, 25*, 343-352.

Barnard, K., Duygulu, P., Forsyth, D., de Freitas, N., Blei, D. M., & Jordan, M. I. (2003). Matching words and pictures. *The Journal of Machine Learning Research, 3*, 1107-1135.

Beck, M. R., Peterson, M. S., Boot, W. R., Vomela, M., & Kramer, A. F. (2006). Explicit memory for rejected distractors during visual search. *Visual Cognition, 14*(2), 150-174.

Becker, M.W., & Rasmussen, I.P. (2008). Guidance of of Attention to Objects and

Locations by Long-Term Memory of Natural Scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(6), 1325–1338.

Blackwood, W. O., Anderson, T. R., Bennett, C. T., Corson, J. R., Endsley, M. R., Hancock, P. A., et al. (1997). *Tactical displays for soldiers: Human factors considerations. Panel on Human Factors in the Design of Tactical Display Systems for the Individual Soldier (panel chair Blackwood), National Research Council*: National Academy Press.

Brabham, D. C. (2008). Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence, 14*(1), 75.

Brady, T. F., Konkle, T., Oliva, A., & Alvarez, G. A. (2008). Detecting changes in real-world objects: The relationship between visual long-term memory and change blindness. *Proc Natl Acad Sci U S A.*

Caird, J. K., Edwards, C. J., Creaser, J. I., & Horrey, W. J. (2005). Older Driver Failures of Attention at Intersections: Using Change Blindness Methods to Assess Turn Decision Accuracy. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *47*(2), 235-249.

Chun, M. M., & Jiang, Y. (1998). Contextual Cueing: Implicit Learning and Memory of Visual Context Guides Spatial Attention. *Cognitive Psychology*, *36*(1), 28-71.

Chun, M. M., & Nakayama, K. (2000). On the Functional Role of Implicit Visual Memory for the Adaptive Deployment of Attention Across Scenes. *Visual Cognition, 7*(1), 65-81.

CNN. (2007, April 16, 2007). Student shot video of campus shooting. from http://www.cnn.com/2007/US/04/16/vtech.witness/index.html

Cucchiara, R. (2005). *Multimedia surveillance systems*. Paper presented at the Proceedings of the third ACM international workshop on Video surveillance & sensor networks.

Czerwinski, M., Tan, D. S., & Robertson, G. G. (2002). *Women take a wider view*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves.

Dee, H. M., & Velastin, S. A. (2008). How close are we to solving the problem of automated visual surveillance?: A review of real-world surveillance, scientific progress and evaluative mechanisms. *Mach. Vision Appl., 19*(5-6), 329-343.

Dey, A. K. (2001). Understanding and using context. *Personal and ubiquitous computing, 5*(1), 4-7.

Michael Dorr. Effects of Gaze-Contingent Stimuli on Eye Movements. Diploma thesis, Universität zu Lübeck, 2004.

Downing, C. J. & Pinker, S. (1985). The spatial structure of visual attention. In M. Posner and O.Marin (Eds.), *Attention and Performance XI: Mechanisms of attention and visual search*. Hillsdale, NJ: Erlbaum.

Endsley, M.R. (2000). Theoretical underpinnings of situation awareness: a critical review. In: Endsley,M.R., Garland, D.J. (Eds.), Situation Awareness Analysis and Measurement. Lawrence Erlbaum,Mahwah, NJ, pp. 3–32.

Michael Dorr. Effects of Gaze-Contingent Stimuli on Eye Movements. Diploma thesis, Universität zu Lübeck, 2004

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychol Rev, 96*(3), 433-458.

El-Nasr, M. S., & Su, Y. (2006). *Visual attention in 3D video games*. Paper presented at the Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology.

facebook. (2009). Facebook | Photos: Viewing, editing, tagging, and privacy. Retrieved June 23, 2009, from http://developers.facebook.com/help.php?page=831

Ferrin, G., Snidaro, L., Canazza, S., & Foresti, G. L. (2008). Soft data issues in fusion of video surveillance, 2008 11th International Conference on. *Proceedings of the 11th Annual International Conference on Information Fusion*, Cologne, Germany

Gagnon, D. (1985). Videogames and spatial skills: An exploratory study. *Educational Technology Research and Development, 33*(4), 263-275.

Geng, J. J., & Behrmann, M. (2005). Spatial Probability as an Attentional Cue in Visual Search. *Perception and Psychophysics, 67*(7), 1252.

Glater, J. D. (2009, January 17, 2009). Helping Keep a City Clean, and Maybe Safer *The New York Times*, from http://www.nytimes.com/2009/01/18/us/18trash.html

google. (2009). Street View. from http://maps.google.com/help/maps/streetview/faq.html

D. Gorodnichy (2006) Ace surveillance: The next generation surveillance for long-term monitoring and activity summarization. Presented at CRV'06 Workshop on Video Processing for Security, Quebec City, Quebec, Canada.

Green, C. S., & Bavelier, D. (2007). Action-video-game experience alters the spatial resolution of vision. *Psychological Science, 18*(1), 88-94.

Green, C. S., & Bavelier, D. (2006). Enumeration versus multiple object tracking: the case of action video game players. *Cognition, 101*(1), 217-245.

Gualdi, G., Prati, A., & Cucchiara, R. (2008). Video Streaming for Mobile Video Surveillance. *Multimedia, IEEE Transactions on*, *10*(6), 1142-1154.

Guerrier, J. H., Manivannan, P., & Nair, S. N. (1999). The role of working memory, field dependence, visual search, and reaction time in the left turn performance of older female drivers. *Applied Ergonomics, 30*(2), 109-119.

Hall, D. L., McNeese, M., Llinas, J., & Mullen, T. (2008). *A framework for dynamic hard/soft fusion.* Paper presented at the Information Fusion, 2008 11th International Conference on Information Fusion, held in Cologne, Germany on July 2008.

Hall, D.L., Aungst, S. (2008, May). *The Use of Soft Sensors and I-Space for Improved Combat* ID. Paper presented at the Air Force Research Laboratory Workshop on Human Factors Issues in Combat Identification, 2008 Apache Junction, AZ

Hayhoe, M. (2000). Vision Using Routines: A Functional Account of Vision. *Visual Cognition, 7*(1-3), 43-64.

Healey, C. G., & Enns, J. T. (1999). Large datasets at a glance: combining textures and colors inscientific visualization. *IEEE Transactions on Visualization and Computer Graphics, 5*(2), 145-167.

Hedeker, D., & Gibbons, R. D. (1994). A Random-Effects Ordinal Regression Model for Multilevel Analysis. *Biometrics, 50*(4), 933-944.

Henderson, J. M., & Hollingworth, A. (1999). High level scene perception. *Annual Review of Psychology, 50*, 243-271.

Hi-Tec Solutions (2007, July 17, 2007). Homeland Security products. Retrieved May 27, 2009, 2009

Hillstrom, A. P., & Yantis, S. (1994). Visual motion and attentional capture. *Perception and Psychophysics, 55*(4), 399-411.

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception and Psychophysics, 57*(6), 787-795.

Holmberg, A. (2007). *Eye tracking and gaming Eye movements in Quake III: Arena.* Unpublished Masters, Lund University.

Horowitz, T., & Wolfe, J. (2003). Memory for rejected distractors in visual search? *Visual Cognition*, *10*(3), 257-298.

Howe, J. (2006). The rise of crowdsourcing. *Wired Magazine*, *14*(6), 176-183.

Intille, S. S., & Bobick, A. F. (2001). Recognizing planned, multiperson action. *Computer Vision and Image Understanding, 81*(3), 414-445.

Irwin, D. E., & Zelinsky, G. J. (2002). Eye Movements and Scene Perception: Memory for Things Observed. *Perception & psychophysics, 64*(6), 882-895.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10-12), 1489-1506.

John, B. E., & Kieras, D. E. (1996). Using GOMS for user interface design and evaluation: Which technique? *ACM Transactions on Computer-Human Interaction, 3*(4), 287-319.

Lamy, D., Bar-Anan, Y., Egeth, H. E., & Carmel, T. (2006). Effects of Top-Down Guidance and Singleton Priming on Visual Search. *Psychonomic Bulletin and Review, 13*(2), 287.

Lamy, D., & Tsal, Y. (2000). Object features, object locations, and object files: Which does selective attention activate and when? *Journal of experimental psychology. Human perception and performance, 26*(4), 1387-1400.

Leonard, C. J., & Egeth, H. E. (2008). Attentional guidance in singleton search: An examination of top-down, bottom-up, and intertrial factors. *Visual Cognition, 16*(8), 1078-1091.

Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., et al. (2008). Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey? *Monthly Notices of the Royal Astronomical Society, 389*(3), 1179-1189.

Lossau, K., & Stotz, A. (2008). Practical frameworks for supporting tactical fusion, *Fusion 2008*. Beaver Hollow, NY.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279.

Maloney, L. T., & Wandell, B. A. (1986). Color constancy: a method for recovering surface spectral reflectance. *Journal of the Optical Society of America A, 3*(1), 29-33.

Makris, D., & Ellis, T. (2002). Path detection in video surveillance. *Image and Vision Computing, 20*(12), 895-903.

Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. *Journal of Vision, 6*(1), 2-8.

Menneer, T., Barrett, D. J. K., Phillips, L., Donnelly, N., & Cave, K. R. (2007). Costs in searching for two targets: dividing search across target types could improve airport security screening. *Applied Cognitive Psychology, 21*(7), 915-932.

Moore, C. M., Egeth, H., Berglan, L., & Luck, S. J. (1996). Are attentional dwell times inconsistent with serial visual search? *Psychonomic Bulletin & Review, 3*, 360-365.

Myles-Worsley, M., Johnston, W. A., & Simons, M. A. (1988). The influence of expertise on X-ray image processing. *J Exp Psychol Learn Mem Cogn, 14*(3), 553-557.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature, 434*(7031), 387-391.

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research, 46*(5), 614-621.

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences, 11*(12), 520-527.

Oliva, A., Torralba, A., Castelhano, M. S., & Henderson, J. M. (2003). *Top-down control of visual attention in object detection.* Paper presented at the Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on.

O'Regan, J. K., Rensink, R. A., & Clark, J. J. (1999). Change-blindness as a result of /`mudsplashes/'. *Nature, 398*(6722), 34-34.

Owsley, C., Ball, K., McGwin, J. G., Sloane, M. E., Roenker, D. L., White, M. F., et al. (1998). Visual Processing Impairment and Risk of Motor Vehicle Crash Among Older Adults (Vol. 279, pp. 1083-1088): Am Med Assoc.

Palen, Leysia, Sarah Vieweg, Jeannette Sutton, Sophia Liu and Amanda Hughes (2007). Crisis Informatics: Studying Crisis in a Networked World. *Proceedings of the Third International Conference on E-Social Science*, Ann Arbor, MI, Oct 7-9, 2007

Peterson, M. S., Kramer, A. F., Wang, R. F., Irwin, D. E., & McCarley, J. S. (2001). Visual search has memory. *Psychological Science, 12*(4), 287-292.

Pravia, M. A., Prasanth, R. K., Arambel, P. O., Sidner, C., & Chee-Yee, C. (2008). *Generation of a fundamental data set for hard/soft information fusion.* Paper presented at the Information Fusion, 2008 11th International Conference on.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink. *Journal of Experimental Psychology: Human Perception and Performance, 18*(3), 849-860.

Rayner, K. (1995). Eye movements and cognitive processes in reading, visual search, and scene perception. *Eye movement research: Mechanisms, processes and applications*, 3-22.

Reingold, E. M., Charness, N., Schultetus, R. S., & Stampe, D. M. (2001). Perceptual automaticity in expert chess players: Parallel encoding of chess relations. *Psychonomic Bulletin & Review, 8*(3), 504-510.

Rousselet, G., Joubert, O., & Fabre-Thorpe, M. (2005). How long to get to the" gist" of real-world natural scenes? *Visual Cognition, 12*(6), 852-877.

Sadasivan, S., Greenstein, J. S., Gramopadhye, A. K., & Duchowski, A. T. (2005). Use of eye movements as feedforward training for a synthetic aircraft inspection task. *Conference on Human Factors in Computing Systems*, 141-149.

Senior, A., Pankanti, S., Hampapur, A., Brown, L., Tian, Y. L., Ekin, A., et al. (2005). Enabling video privacy through computer vision. *IEEE Security & Privacy, 3*(3), 50-57.

Shih, S. I., & Sperling, G. (1996). Is there feature-based attentional selection in visual search? *J Exp Psychol Hum Percept Perform, 22*(3), 758-779.

Shimozaki, S. S., Eckstein, M. P., & Abbey, C. K. (2005). Spatial profiles of local and nonlocal effects upon contrast detection/discrimination from classification images. *Journal of Vision, 5*(1), 45-57.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception-London-, 28*, 1059-1074.

Simons, D. J., Franconeri, S. L., & Reimer, R. L. (2000). Change blindness in the absence of a visual disruption. *Perception-London-, 29*(10), 1143-1154.

Simons, D. J., & Levin, D. T. (1998). Brief Reports: Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review, 5*(4), 644-649.

Simons, D. J., Rensink, R. A., Hayhoe, M., Chun, M. M., Nakayama, K., Shapiro, K. L., et al. (2000). Current Approaches to Change Blindness. *Change Blindness and Visual Memory, 2000*(7), 1-15.

Souto, D., & Kerzel, D. (2009). Involuntary cueing effects during smooth pursuit: facilitation and inhibition of return in oculocentric coordinates. *Experimental Brain Research, 192*(1), 25-31.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643-662.

Tatler, B. W., & Melcher, D. (2007). Pictures in mind: Initial encoding of object properties varies with the realism of the scene stimulus. *Perception - London, 36*(12), 1715.

Thorne, D. R., Johnson, D. E., Redmond, D. P., Sing, H. C., Belenky, G., & Shapiro, J. M. (2005). The Walter Reed palm-held psychomotor vigilance test. *Behavior Research Methods, 37*(1), 111-118.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual Guidance of Eye Movements and Attention in Real-World Scenes: The Role of Global Features in Object Search. *Psychological Review, 113*(4), 766-786.

Troscianko, T., Holmes, A., Stillman, J., Mirmehdi, M., Wright, D., & Wilson, A. (2004). What happens next? The predictability of natural behaviour viewed through CCTV cameras. *Perception, 33*(1), 87-101.

Turso, V. A., & Duckett, T. (2004). Sanitation to Appoint 410 New Sanitation Workers, Promote 23 Superintendents, Honor 'Employees of the Year' and Acts of Bravery. from http://www.nyc.gov/html/dsny/html/pr2004/062504.shtml

Want, R., & Schilit, B. (2001). Expanding the horizons of location-aware computing. *Computer, 34*(8), 31-34.

Werner, S., & Thies, B. (2000). Is" Change Blindness" Attenuated by Domain-specific Expertise? An Expert-Novices Comparison of Change Detection in Football Images. *Visual Cognition*, *7*(1), 163-173.

Wesensten, N. J., Belenky, G., & Balkin, T. J. (2005). Cognitive readiness in network-centric operations. *Parameters, 1*(35), 94-105.

Wickens, C. D., Fadden, S., Merwin, D., & Ververs, P. M. (1998). Cognitive factors in aviation display design. *Proceedings of the 17th Digital Avionics Systems Conference,* Institute of Electrical & Electronic Engineers.

Wickens, C. D., & Long, J. (1995). Object versus space-based models of visual attention: Implications for the design of head-up displays. *Journal of Experimental Psychology: Applied, 1*(3), 179-193.

Wolfe, J. M. (2003). Moving towards solutions to some enduring controversies in visual search. *Trends in Cognitive Sciences, 7*(2), 70-76.

Xu, K., Zhu, M., Zhang, D., & Gu, T. (2008). *Context-aware content filtering presentation for pervasive & mobile information systems*. Paper presented at the Proceedings of the 1st international conference on Ambient media and systems.

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance, 10*(5), 601-621.

Yarbus, A. L. (1967). Eye movements during perception of complex objects. *Eye movements and vision, 7*, 171-196.

Yeh, M., Wickens, C. D., & Seagull, F. J. (1999). Target Cuing in Visual Search: The Effects of Conformality and Display Location on the Allocation of Visual Attention. *Human Factors, 41*(4).

Appendix A

Informed Consent Form for Social Science Research
The Pennsylvania State University

ORP OFFICE USE ONLY
DO NOT REMOVE OR MODIFY
IRB#**28457 Doc. #1**
The Pennsylvania State University
Office for Research Protections
Approval Date: 02-26-09 LSY
Expiration Date: 05/18/2009 LSY
Social Science Institutional Review Board

Title of Project: Technological Approaches for Reporting of Information

Principal Investigator: Alice Shapiro
ars301@psu.edu
0329B Information Science and Technology Building
University Park, PA 16802
(814) 865 - 6847

Advisor: Dr. David Hall
dhall@ist.psu.edu
316 B IST Building, University Park, PA 16802
Telephone: 814-867-2154

Other Investigator(s):

Bimal Balakrishnan bub106@psu.edu
Varun Adibhatla vxa136@psu.edu
Mark S. Pfaff msp100@psu.edu
Michael McNeese MMcNeese@ist.psu.edu

USE Lab - 314 IST Building, University Park, PA 16802
Telephone 814-865-6847

1. **Purpose of the Study:** The study in which you will be participating is designed to study the effect of different types of technology on information elicitation and reporting.

2. **Procedures to be followed:** If you agree to take part in this research, you will be asked to take part in experimental sessions lasting ONE hour. You will be asked to take a survey asking you to answer questions based on how you use mobile technology. You may be asked to do one or all of the following. You then may be asked to perform a text messaging training task. You will then watch a video, and submit responses during the video using a mobile phone, verbal response, mouse, keyboard, or similar personal computing equipment. You will then be asked to fill out a survey regarding your experience on paper or online after the experiment. If you are part of a focus group, you will participate with other participants and engage in a discussion initiated and moderated by the researcher. If you are being interviewed individually, the researcher may ask you specific questions relevant to the context of this research.

3. **Discomforts and Risks:** There are no risks in participating in this research beyond those experienced in everyday life. If anything does cause discomfort or harm to you, you are free to stop your participation in this study at any point by informing the experimenter.

4. **Benefits:** The benefits to you include learning about different forms of electronic communication and/or visual search and reporting of information. Furthermore, the benefits of this study to society include a better understanding of the best ways to present visual information and elicit feedback electronically.

5. **Duration/Time:** Your participation in this research will take no longer than one hour.

6. S**tatement of Confidentiality:** Your participation in this research is confidential. In the event of publication of this research, no personally identifying information will be disclosed. Confidentiality of the results of this study, as they apply to subjects in general, will be maintained as no names or identification of any sort will be tagged to the data. The survey does not ask for any information that would identify who the responses belong to. Therefore, your responses are recorded confidentially. The following may review and copy records related to this research: The Office of Human Research Protections in the U.S. Dept. of Health and Human Services; The Penn State University Social Science Institutional Review Board (IRB); The Penn State University Office for Research Protections. Confidentiality will be maintained to the degree permitted by the technology used. Specifically, due to the current state of the art of Internet technology security, no guarantees can be made regarding the interception of data sent via the Internet by any third parties. If this research is published, no information that would identify you will be written since your name is in no way linked to your responses. If you speak about the contents of the focus group outside the group, it is expected that you will not tell others what individual participants said

7. **Right to Ask Questions:** Please contact Alice Shapiro ars301@psu.edu with questions, complaints or concerns about this research. You can also contact this individual if you feel this study has harmed you. You may ask any questions about the research procedures during the experiment. After you have finished participating, you will receive a more detailed explanation of the study. Any questions you have at this time will be answered. Questions about your rights as a research participant may be directed to Penn State University's Office for Research Protections at (814) 865-1775.

8. **Payment for participation: Payment for participation:** In return for participation in this study, you will be entered into a drawing for a $20 gift certificate at Starbucks Coffee. You may also receive 10 extra credit points (out of approximately 280 points) if you are enrolled in the course IST 440W (with Dr. Hall), or have an additional 1% added to your final grade if you are enrolled in COMM 420 (with Instructor Qian Xu. If you choose not to engage in this experiment, an alternative assignment will be available for you. You will be asked to read an article relevant to the experiment. Please contact the principal investigator using the information at the top of the page for additional information about alternative to extra credit.

9. **Voluntary Participation:** Your decision to be in this research is voluntary. You can stop at any time. You do not have to answer any questions you do not want to answer. Refusal to take part in or withdrawing from this study will involve no penalty or loss of benefits you would receive otherwise.

You must be 18 years of age or older to consent to take part in this research study. If you agree to take part in this research study and the information outlined above, please sign your name and indicate the date below.

You will be given a copy of this consent form for your records.

____ ____________________
Participant Signature Date

____ ____________________
Person Obtaining Consent Date

Experimenter fills out this area

Sequence __1_

Number ______

Appendix B Questionnaire

Age ________

Gender ________

Clip 1

1. Did you locate the target?

2. Did you recognize the location of the crowd?

3. Did you recognize any individuals within the crowd?

Number ______

1. Did you find the size or other features of the attention directing tool to be distracting?

2. Have you ever worked for campus safety or any other security related service?
   If yes, what type? _______________
   For approximately how long? (months and years) __________

3. Have you ever worked for law enforcement of any type?
   If yes, what type? ________________
   For approximately how long (months and years)? ___________

4. Have you ever worked for any kind of emergency services department?
   If yes, what type? _________________
   For approximately how long (months and years)? ____________

5. Can you lipread? (circle response)
   Yes
   No
   Some
   About as much as the average person