Where's Waldo

Studying Object Recognition in Crowded Pictures Through Gaze Analytics

Hannah Schilling
Clemson University
Clemson, South Carolina
shschil@g.clemson.edu

Joseph Flowers
Clemson University
Clemson, South Carolina
jhflowe@g.clemson.edu

Joshua Glenn
Clemson University
Clemson, South Carolina
jaglenn@g.clemson.edu

Josue Figueroa
Clemson University
Clemson, South Carolina
josuef@g.clemson.edu

ABSTRACT

When people are locating an object, they tend to use a variety of methods to scan the image that they are searching through. People will use systematic searching, going by rows or columns, or randomly look throughout the picture. Here, we will analyze how overlaying a grid over images will affect participant's ability to locate a given object in a crowded image. In our experiment, we used a Gazepoint GP3 Desktop Eye-Tracker to track and time 17 participants finding of Waldo across 2 different experiments using 2 stimuli for each. One experiment featured 2 "Where's Waldo" images with a grid overlayed on top of it and the other experiments was same 2 images, but without the grid. The statistical analysis of the recorded data did not yield any significant differences due to several participant's data being invalid, all who wore corrective lenses, but we were able to do an analysis of our own by observing the raw data and their averages. We found that the images without the grid had much faster average mean times for finding Waldo, while the images with the grid had much slower mean times for finding Waldo. The errors caused by loss of data and failure to correctly calibrate participants with corrective lenses, as well as the lack of more variety of stimuli have limited this experiment. These findings can be corrected and expanded upon in further studies to find the effectiveness of a systematic search using a grid versus not using one.

CCS CONCEPTS

• **Computing methodologies** → **Visual inspection**; *Shape inference*; *Object detection*;

KEYWORDS

eye tracking, object detection, visual inspection

ACM Reference Format:
Hannah Schilling, Joseph Flowers, Joshua Glenn, and Josue Figueroa. 2018. Where's Waldo: Studying Object Recognition in Crowded Pictures Through Gaze Analytics. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

There are many reasons today to study the search patterns of humans that go far beyond the scope of basic research. Studying how people search through images to locate desired objects can help improve the way that information is presented to people, such as search results, or how information on a page is laid out. One facet of this school of thought is how to make it easier for people to find the information or object that they are looking for. The experiment outlined by this paper is focused on how overlaying a grid can facilitate a user's ability to locate a given object within a crowded image.

1.1 Goals

Through this study we hope to understand how different distractions affect a person's ability to search for objects within an image. By overlaying grids of differing size onto an image, we hope to find a correlation between the size of the grid squares with how fast a user can locate a desired object within a crowded image.

1.2 Motivation

We wish to study how difficult it is for participants to find objects in a crowded image, and determine if there is a way to make visually searching images easier and more efficient. We hypothesize that overlaying a grid onto an image makes it easier to located specified objects, and that the size of the grid squares will affect the user's ability to locate objects.

1.3 General Hypothesis

We hypothesize that overlaying a grid will affect the participants' ability to visually locate Waldo. Given a grid, participants will be able to locate Waldo more quickly than participants just presented with the same images, without a grid.

2 BACKGROUND

Previous eye tracking research incorporating "Where's Waldo?" has been done in order to find a variety of different things. One study being "Finding Waldo: Learning about Users from their Interactions". In this study, machine learning techniques are used to

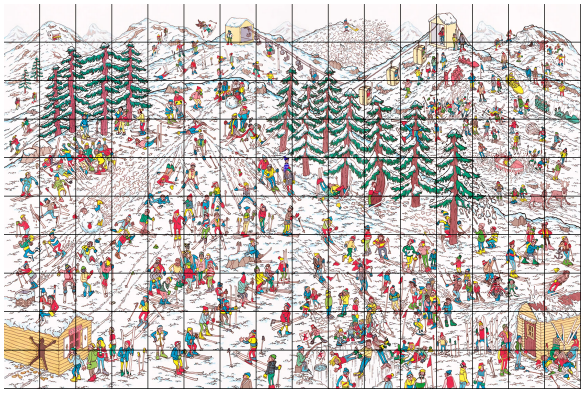

Figure 1: The first graphic that was presented to the experimental group.

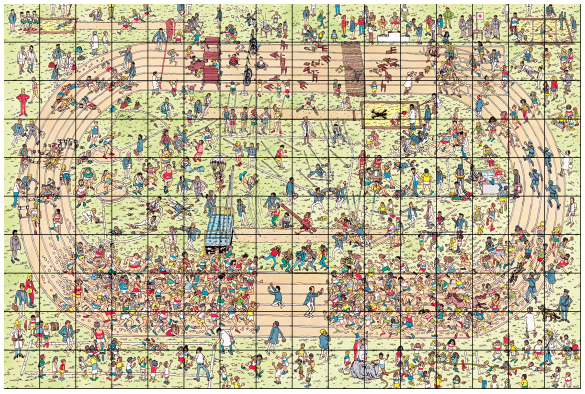

Figure 2: The second graphic that was presented to the experimental group.

Figure 3: The first graphic that was presented to the control group.

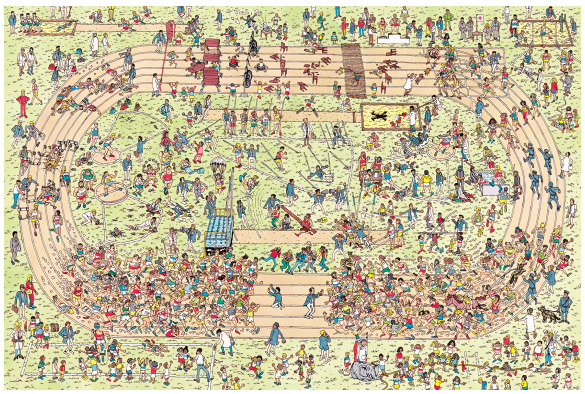

Figure 4: The second graphic that was presented to the control group.

predict the speed at which the user's are able to find Waldo as well as inferring user personality traits [1]. Another study, which went unpublished, was conducted to learn of the correlation between the reading of natural language (e.g left to right, from top to bottom) and finding Waldo in a "Where's Waldo?" game, specifically using the correlation to the user's words per minute in relation to the amount of time they fixated on a specific part of the screen [2]. The hypothesis was that users with a higher reading speed would have shorter fixation times, although this was not proven true based on the results. This does however relate to our study in that we are testing the effect of a systematic search, using a grid, versus not using one to discover whether humans are able to find Waldo with more ease and speed if the image is split into sections, searching through it almost as one would read natural language.

3 EXPERIMENT METHODOLOGY

3.1 Apparatus

The monitor to be used for the eye tracking experiment is a Dell P2213 22" LED LCD monitor with a resolution of 1680 x 1050. Eye movements will be tracked using a GP3 Desktop Eye-Tracker designed by GazePoint. The sampling rate of the participant's eyes will be 60Hz with a visual angle accuracy of 0.5-1 degrees. The GP3 tracks participant's eye movements, fixations, and the distance between the participant's eyes and screen.

3.2 Stimulus

The main stimuli of the experiment are 2 "Where's Waldo?" images with Waldo hidden within the busy image (Figures 1-4). Participants were directed to look at an image with a dot in the center to focus on before proceeding to the "Where's Waldo" image. Once Waldo was found within the busy image, the participant were directed to another image with a dot to regain their focus before advancing to the next image. In order to rule out dishonesty by the participant about finding Waldo, the researchers asked the participant to indicate where they found Waldo, using their mouse to point where he is located.

3.3 Subjects

The study consisted of 18 participants. The participants were undergraduate students from Clemson University and ranged from people with perfect vision to people with corrected vision. The participants were college-age students with ages ranging from 18-25.

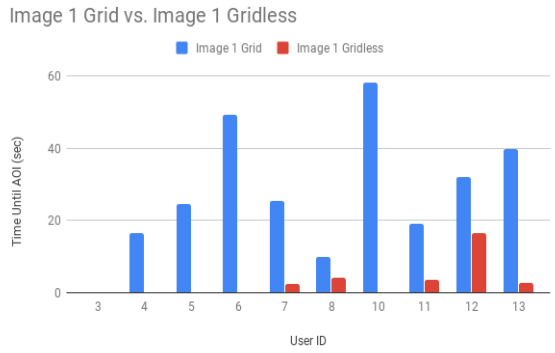

Figure 5: Graph showing the time it took for subjects to find Waldo in Image 1, with grid vs without grid.

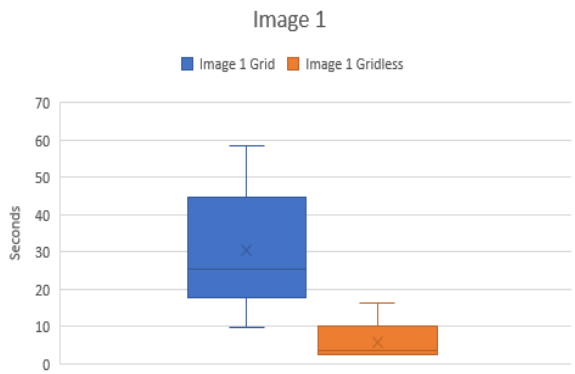

Figure 6: Graph showing the mean and median time it took for subjects to find Waldo in Image 1, with grid vs without grid.

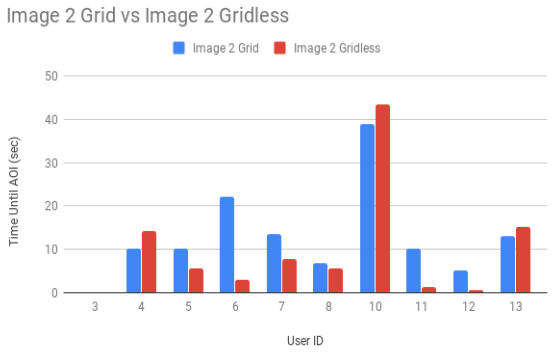

Figure 7: Graph showing the time it took for subjects to find Waldo in Image 2, with grid vs without.

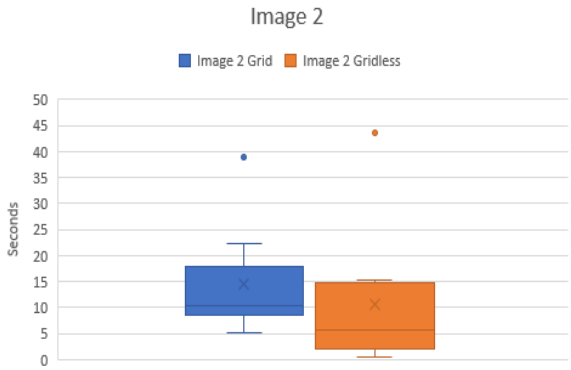

Figure 8: Graph showing the mean and median time it took for subjects to find Waldo in Image 2, with grid vs without.

3.4 Experimental Design

The type of experimental design used in the investigation was a within-subjects design. Participants will be provided a "Where's Waldo?" image and asked to identify Waldo in an unlimited time frame using GazePoint software to track the time it takes to find Waldo. Once the participant finds Waldo, they will indicate where they have found him using the mouse pointer and be introduced to a blank image to restart the process all over again. If part of the control group, the images are presented unaltered. If part of the experimental group, the same images from the control group will be used, but this time with a grid overlaying the images. Each participant will look at the stimulus in random order for unbiased purposes and to reduce fatigue. The independent variables are the images overlaid with a grid. The dependent variable is the fixation speed and time being searched for Waldo. The GazePoint software will track the length of time each participant takes to find Waldo.

Participants will first be introduced to the researchers in an environment-friendly computer lab. The participants will then be asked to be seated near one of the monitors with the eye tracker. Once seated, participants will be given an informational letter which includes risks involved with the experiment, potential benefits, their confidentiality, our contact information, and lastly let the participant know that this is a voluntary participation and they are welcomed to opt out of the study. Once the participant has provided consent, our next step will be to instruct what to do next in a verbal manner.

First, the participant will be asked questions regarding demographics such as age, gender, and occupation. The participant will then begin by fixating on the targets given on the screen to first calibrate the GazePoint software. Following calibration, the participant will be introduced to a screen containing the exact Waldo that they will be finding in the following scenes. If the participant is part of the experimental group, they will be presented with a grid that has arrows indicating how the participant will be searching through the imagine in order to find Waldo. If the participant is part of the control group, this step is skipped. Next, the participant will be presented with a blank image and a dot and then the screen was switched to the test stimuli and instructed to find Waldo. The participant will be timed on how long it took them to find Waldo during each stimulus. Once he was found, the participant will notify the researcher and will move the mouse to Waldo so the researcher is sure he has been found. After the first stimuli, the researcher

will present the participant with the blank image again. Then, the second stimuli will be shown until the participant finds Waldo.

4 RESULTS

Based on the data collected from the experiments, The gridless images had faster times until The first fixation of the AOI that was placed over Waldo. The average time to find Waldo in the images without the grid was 5.825 for image 1 and 10.733 for image 2. the average time for finding Waldo in the Grid images was 30.55 for image 1 and 14.483 for image 2. To find the effectiveness of using a grid vs. not using one, we used a Two-Way ANOVA test in R for the average time of image1, testing image 1 with a grid against image 1 without. This was also done for image 2. The results of image 1 yielded a P-Value of 0.00641 while the results of image 2 yielded a P-Value of 0.515. The significance level of image 2 are much higher than that of image 1 most likely due to a loss of data for 4 subjects in the gridless experiment for image 1 and a loss of data for 1 subject in the grid experiment for image 1. Due to these errors in the experiment, the data collected doesn't serve as significant data to support or go against our hypothesis.

The raw data however does serve as a good source of evidence to test our hypothesis. The data shows that the average times of the gridless approach were much faster than those of the grid approach. This is most likely due to a problem with our experiment where images with the grid gave the users a large constraint, slowing their possible finding time. We however did observe that users who took the grid experiment found Waldo faster on the second image than they did on the first image, while users who took the gridless experiment found Waldo faster on the first image than the second image. This could possibly be due to the those specific images as image 1 has a mostly white background, so Waldo is more apparent, while he blends into the background a bit more in image 2 and it's a bit more crowded. The grid seems to serve as a guide to pinpoint him faster than the user would otherwise without it. Despite this, the grid still has the issue of it constraining the users traversal through the image.

5 DISCUSSION

Due to time constraints of the experiment, only 18 out of the anticipated 20 participants were put through the experiment. Out of the collected data, one participant was excluded due to the software not detecting their pupils through the corrective lenses that the participant was wearing. This participant was put through the gridless experiment. Three participants did not have data for the first gridless image that they were presented; the data for these participants will be discussed with respect to the second image only.

The stimuli was not randomized, each participant was shown the same images in the same order, in order to facilitate the comparison of gathered data between the grid experiment and the gridless experiment.

5.1 Further Research

In further iterations of this experiment, we would consider increasing the stimuli count and randomizing the order of images seen by the participant. In addition to increasing the stimuli count, we would present a wider variety of "Where's Waldo?" puzzles, taking into account Waldo's location, the amount of other characters in the picture, as well as the background color that could serve to make him less or more visible. This addition could also possibly circumvent the issue of the grid experiment having a much larger mean finding time as images that could benefit from the inclusion of a grid could be used, giving more accurate data and results. We could also let the user who participate in the grid experiment start from any on the grid they want rather than beginning in one of the top corners, or even allow them to freely traverse the image, just as participants of the gridless experiment would, using the grid as a guide rather than a path.

REFERENCES

[1] Brown, E. T., Ottley, A., Zhao, H., Lin, Q., Souvenir, R., Endert, A., Chang, R. (2013, October). Finding Waldo: Learning about Users from their Interactions. IEEE.

[2] Andrews D., Hill E., Lewis H., Martin M. (2017). Analysis of Words Per Minute Correlated to Fixation Length While Searching. In Proceedings of CPSC4120. ACM, New York, NY, USA, 5 pages.