Using Image Saliency and Regions of Interest to Encourage Stronger Graphical Passwords

Mohammad N. Alshehri *
Information Technology Department
Institute of Public Administration
Riyadh, Saudi Arabia
shehrimo@ipa.edu.sa

Heather Crawford
Harris Institute for Assured Information
Florida Institute of Technology
Melbourne, FL, 32940
hcrawford@fit.edu

ABSTRACT

A graphical password guiding image serves as a visual prompt to improve password memorability. However, passwords may be easily guessed if the guiding image contains hotspots, or commonly chosen (e.g., 'clickable') points that are predictable via automated means. In this paper, we propose a method to determine graphical password guiding image suitability in terms of potential password strength. Our method uses image saliency to measure image suitability; the higher the saliency, the more suitable the image. Next, we evaluate the regions of interest (e.g., circles, faces, corners, etc.) of suitable images to predict the strength of resultant graphical passwords. We provide support for our method in two ways: first, we analyzed the guiding images and resulting graphical password strength from an existing dataset and secondly, we conducted our own user study to measure the usability and memorability of the same guiding images in terms of registration, login and recall times. We found that the more visually salient the image, the stronger the resulting graphical passwords in terms of entropy with little or no effect on usability and memorability. Furthermore, users tended to select more suitable images even when given the choice of less suitable images. Thus, our approach may be used to improve the strength of graphical passwords before the user chooses a single point or action simply by excluding unsuitable guiding images.

CCS Concepts

•Security and privacy → Graphical / visual passwords; Usability in security and privacy;

Keywords

Authentication; Graphical Passwords; Usable Security

*This author was a student at Florida Institute of Technology at the time this research was conducted.

ACSAC '16 December 5 – 9, 2016, Los Angeles, CA, USA

ISBN 978-1-4503-4771-6/16/12…$15.00

DOI: http://dx.doi.org/10.1145/2991079.2991108

1. INTRODUCTION

Smartphones have become a fully-fledged computing environment that many people depend upon. They are used to store and access personal information such as medical records, banking information, private email and text messages, even when they were deleted [15]. Many smartphones use a soft keyboard, which makes typing passwords more error prone compared to a traditional keyboard [2, 26]. In addition, password composition rules are optimized for desktop keyboards since they require the use of special characters, which are more effortful to type on a soft keyboard [17]. Since many smartphones owners use traditional text-based passwords that are typed via soft keyboards, typing errors and slow entry rate may negatively affect the usability of textual passwords on these platforms [28] while creating easier to type passwords may negatively affect their strength.

Graphical passwords do not rely on traditional textual entry, instead depending on selecting a series of points on one or more images. Despite the natural match between touch-based mobile devices and graphical passwords, few mobile manufacturers have used graphical passwords for authentication. One reason may be that click-based graphical passwords are prone to *hotspots*, which are frequently chosen points on the image that draw the user's visual attention. Hotspots are chosen frequently because they are more memorable than other click points, which makes the password easier to remember [32], but also increase the likelihood that a dictionary-based attack will succeed [32]. This is akin to having knowledge of a target before attacking their textual password; the attacker uses this knowledge to guess more likely passwords first. Advice by researchers with regards to hotspots is to either choose images without hotspots or to choose other, less likely points [5]. However, this advice may lead to less memorable graphical passwords, which may affect usability and user adoption. Instead, we propose that hotspots are a positive part of graphical passwords since an image with many hotspots[1] also has many possible passwords, and that this can be leveraged to create graphical passwords that are both strong and memorable.

In this paper we present a method for improving the strength of graphical passwords while taking advantage of the memorability of visually salient regions. Our approach,

[1]We use the term *hotspot* here to conform to the terminology commonly used in graphical password research, as it generally refers to points on an image that are frequently chosen in graphical passwords. We make a distinction between a frequently chosen point (hotspot) and a visually salient point (RoI) since the latter may or may not be selected frequently.

which begins with measuring the visual saliency of the image and then maps the salient regions to objects that appear in the guiding image, has the potential to remove images that do not contain enough salient regions to make dictionary-based attacks less successful. By allowing users to choose guiding images that are sufficiently complex, we improve password strength before the user makes a single click. We tested our method on thousands of graphical passwords from the dataset by Zhao *et al.* [37] and found a strong positive correlation between image saliency, number and type of regions of interest and the resultant theoretical and practical password spaces. We supported this result by conducting a user study that used the same guiding images as Zhao *et al.* to study the usability and memorability aspects. Our results showed that participants found the speed and ease of using a graphical password acceptable, and that most participants were able to remember their graphical password over time. Finally, we found that the chosen click points were spread over the detected regions of interest, showing that practical password space is not dramatically reduced when user choice is taken into account.

2. BACKGROUND AND RELATED WORK

2.1 Effect of Image Content on User Password Selection

Prediction resistance is a key trait for graphical password security. Selecting an image with predictable hotspots may result in passwords that are easily guessed via dictionary attacks. Stobert *et al.* [30] studied the effect of guiding images on the frequency of creating the same pattern of passwords. They found that users tend to follow patterns when creating graphical passwords, whether using a blank background or guiding image, and that these patterns tended to result in weaker passwords. Davis *et al.* [9] used face and story schemes to study the impact of images on user choice. They found that faces, particularly those of the same race and/or gender as the user, were more appealing. Their results also agree with those of Stobert *et al.*: users in their study followed predictable patterns when creating their graphical passwords.

Dunphy and Yan [12] explored the impact of using a background image on the security of Draw-A-Secret graphical passwords. Their results were that users who were shown background images created more complex and more memorable passwords than those who were not shown background images. Moreover, background images participated in reducing the probability of creating predictable passwords that followed known patterns. Similarly, Golofit [16] found that images containing recognizable places are the most frequently selected areas when creating graphical passwords. These results show that the guiding image has an effect on the areas of the image selected when users create graphical passwords

2.2 Hotspot Detection

Avoiding hotspots may limit graphical password predictability [7, 8, 32]. Consequently, implementations have been proposed to encourage users to avoid hotspots when creating graphical passwords. Chaisson *et al.* [7] introduced Persuasive Cued Click Points (PCCP), an implementation that guided users to click within randomly chosen regions determined by the system rather than on hotspots. The results indicated that this approach helped users to create less guessable graphical passwords, but they did not examine whether the total number of passwords in hotspot-heavy images was sufficient to make this guessing attack viable. PCCP takes advantage of less frequently chosen visually salient regions in some cases, similar to the intuition that guides our work, although we extend this work to show that increasing visually salient features is a positive aspect for graphical passwords in terms of resultant password strength.

Several methods have been proposed to detect image hotspots. Salehi-Abari *et al.* [27] used the Itti algorithm [20] as a bottom-up visual attention method to explore the guessability of graphical passwords. The experiment studied whether users' choice of click points that fall inside detected regions are predictable. Their results indicated that some patterns of graphical passwords, such as diagonal lines, were predicted with reasonable accuracy. van Oorschot *et al.* [33] built a dictionary attack using the Itti visual attention algorithm. The result introduced a significant improvement, compared to [32], for purely automated guessing click-based graphical passwords. The dictionary diagonal lines found over 48% of user passwords for each of the two images used.

Objects that appear in guiding images, such as faces, lines and circles, can be related to the position and prevalence of hotspots. Mayron compared the performance of three different visual attention models, Itti-Koch-Niebur, Graph-Based Visual Saliency (GBVS), and Image Signature to detect the most salient image regions [21]. The result was that these visual attention methods produced good results for detecting objects in images, but Itti-Koch-Neibur and GBVS models performed the best. Mayron and Alshehri evaluated these three models of visual attention to predict the click points used when creating graphical passwords [22]. Their results indicated that, in general, GBVS had the best performance towards predicting graphical passwords.

We believe that hotspots are a benefit to the strength and memorability of graphical passwords provided they are available in sufficient numbers that a dictionary attack is infeasible.

2.3 Measuring Guiding Image Suitability

Image complexity as it applies to graphical passwords is an active research area. Schaub *et al.* [29] explored the design space of graphical passwords on smartphones. They calculated the minimum password length required by each scheme to achieve 14 bit (equivalent to strength of four-digit PIN) and 42 bit (equivalent to strength of seven-character textual password) strength under the assumption of equiprobable password distribution. Their results showed that the studied graphical password implementations required fewer clicks than entering a PIN of comparable strength.

Dirik *et al.* [11] created a model to measure the suitability of background image for the PassPoints graphical password system. They used the color feature to compute the suitability of both complex and simple images. The results indicated that, depending on the model, a dictionary attack was able to predict about 80% of the click points of the simple images and about 70% of click points for the complex images. Moreover, the resultant password entropy was higher in complex compared to simple images, which may aid users in selecting the most appropriate image to use in authentication.

Our work expands that of Dirik *et al.* [11] in several ways. First, we begin with a simple method for determining the

least suitable images in a set and remove those before moving onto a more complex method that takes into account the objects that appear in the guiding image. This first pass simply and easily removes the images that are more likely to be used to create weaker graphical passwords; by not showing these to users we improve the strength of resulting passwords with very little effort. We then use different features for assessing images when compared to Dirik *et al.*: we use image objects such as circles, lines and faces, where Dirik *et al.* combined color contrast, luminosity contrast and foreground versus background objects into a focus of attention map. While Dirik *et al.* rightly limit the number of features they use in their attention map, there are other features that may be considered such as the size of objects, shape, and object category; we therefore extend their work by considering features beyond those they used and determining their effect on the resultant password's entropy. We also extend their work by considering several more images (15 as opposed to their 2) and use two sets of data to confirm our model and results: the dataset of Zhao *et al.* [37], which has over 10,000 passwords from 800 subjects and our own user study dataset with 33 users and 33 passwords. We consider both theoretical and practical password spaces rather than just theoretical. Theoretical password space is the upper bound on the number of possible passwords given a length and a set of possible password click points; it is often an overly generous value that does not take into account user choice. Practical password space (also called effective password space), on the other hand, takes into account the idea that users will prefer certain click points over others and adjusts the calculation of the possible number of passwords accordingly. Finally, we also consider usability and memorability in addition to password strength since strong passwords have been known to suffer from usability and memorability issues that have limited their adoption [28, 36].

Zhao *et al.* [37] identified positions of interest for a set of images based on observing user behavior. They determined the positions of interest by and algorithms to detect image objects, then created a list of generated points. The experiment indicated that faces were the most selected regions for about 60.3% of the passwords' actions that were created on the images. In our work, we extended Zhao *et al.*'s work by considering other regions of interest such as the image corners. Moreover, we considered the diversity of regions of interest. We generated a list of fixed-size ($19x19$) of regions of interest instead of specific points. These regions were prioritized based on the likelihood to be chosen. Our work generated a model that can be used to evaluate the suitability of an image for graphical passwords given the number of regions of interest and their types. This model will simplify the process of choosing an image that may improve the strength of graphical passwords.

2.4 Measuring Password Strength

Theoretical password space is used to measure password strength, including that of graphical passwords. In text-based passwords, the character set from which users select and the minimum password length are two factors that influence password strength. Several types of measures have been proposed, including entropy checkers [24] and comparisons to dictionary or other word lists [10]. Such measurements have been extended to graphical passwords, where the set of characters is equivalent to the set of click points, and password length is equivalent to the number of click points chosen [29, 31].

A strength measurement for recognition-based graphical passwords has been proposed that considers attacks such as guessability, observability, and recordability [13]. A score is computed that represents the security level of the system in terms of resisting the attack. This model does not consider *password* strength; instead, it measures the strength of the system.

Our work builds upon existing research by taking advantage of image saliency as it relates to hotspots to determine whether there is a relationship between image saliency and both the theoretical and practical password spaces. We use this relationship to remove those images with the least amount of saliency (and thus likely having the smallest theoretical password space) as guiding images to encourage stronger passwords without additional user effort.

Our contribution is to provide an approach to selecting suitable images for use with graphical passwords that considers the practical (not theoretical) password space. We show that the practical password space of complex images is higher than for less complex images, and thus we recommend that users and developers take advantage of images with many hotspots rather than avoiding them.

3. OUR APPROACH

We present a two–stage approach to measuring graphical password guiding image suitability designed to increase graphical password strength. Stage 1 measures guiding image suitability based on the salient regions as detected by the GBVS algorithm [18, 19]. The salient regions serve as input to entropy calculations, which are used to measure the theoretical and practical password space for the image. We then relate saliency to image suitability, and propose a decision model that can be used to exclude unsuitable guiding images. Once we have excluded the least suitable images, we further examine what objects in the image comprise the salient regions in Stage 2 of our approach. We detect objects by segmenting the image into regions of $19x19$ pixels, which are then labeled with detected object types including faces, circles, and generic objects. We then compute the likelihood that a user will select the detected region, and then use this likelihood to compute the entropy of the resulting passwords. We evaluate our model using real graphical passwords chosen by users in a study by Zhao *et al.* [37], and also with the results of a user study designed to test memorability and usability.

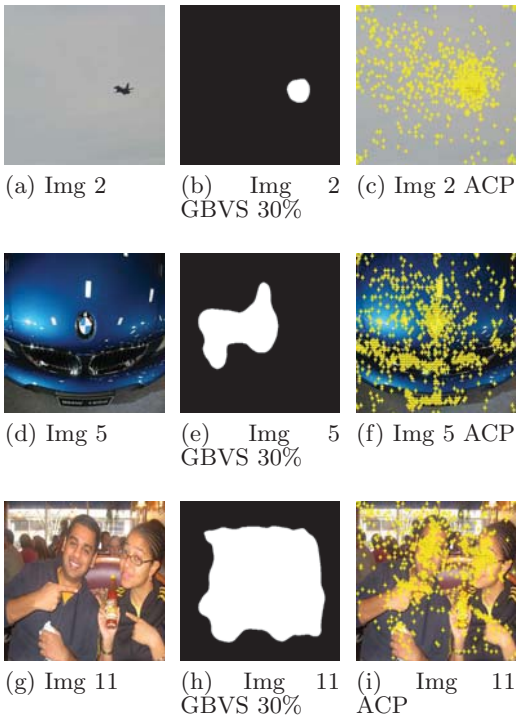
3.1 Data

Training data for both stages consisted of the images and click points provided by Zhao *et al.* [37]. They selected 15 images originally from the PASCAL Visual Object Classes (VOC) Challenge 2007 dataset [14]. The images (see Figure 1 for samples) contain a variety of scenes with varying complexity. Zhao *et al.* collected more than 10,000 passwords from 762 subjects that we use as actual click points (ACP).

3.2 Stage 1: Visual Saliency Determination

We computed the saliency maps of our training images using the MATLAB implementation of GBVS [18, 19]. The generated saliency maps were converted to binary by comparing gray level values of individual pixels to different

Figure 1: Sample of the fifteen images of the training dataset from [37], including computed saliency maps and ACP. Original images from [14].



(a) Img 2 (b) Img 2 GBVS 30% (c) Img 2 ACP

(d) Img 5 (e) Img 5 GBVS 30% (f) Img 5 ACP

(g) Img 11 (h) Img 11 GBVS 30% (i) Img 11 ACP

thresholds ranging from 0% to 90% of the full image saliency. We used a threshold of 30% based on work by Mayron [21] that showed that thresholds greater than 30% provided fewer predictions and thresholds lower than 30% did not provide more predictions. Figures 2b, 2e and 2h show samples of computed saliency maps at the 30% threshold. It can be seen from these figures that the more visually complex images (e.g., Image 11) have a higher proportion of the image that is considered salient.

3.2.1 Image Suitability Measurement

To support the idea that using guiding images with less saliency results in weaker passwords, we measured the entropy of passwords created using the 15 images in Zhao *et al.*'s original study. Our image suitability measurements follow an intuition similar to that of Wang *et al.* [35]: that a higher per-click-point entropy relates to the strength of a given image in providing click points suitable for a strong graphical password. A higher entropy for a click point means a higher theoretical password space, which means the resultant passwords are harder to guess or otherwise crack. Entropy is defined as measuring the amount of uncertainty in the composition of a password [4] and is measured in bits. For example, for a set of characters b and password length k, the entropy of the standard textual password is b^k bits. For a graphical password, characters can be related to salient regions, so the entropy is b^k bits where b is the set of detected salient regions and k is the number of click points chosen for the password. We segmented the salient regions of each image into squares of $19x19$ pixels (and discarded non-salient regions) and counted the number of regions that contained ACP. We chose $19x19$ as a size of a region because it represents the allowed size of the tolerance area for a finger-selected click point that is used in other studies [32, 27]. Table 1 shows the total number of detected salient regions and the proportion of those salient regions that were selected by users. This shows that users do choose salient regions frequently when creating graphical passwords, which provides support for discounting images with small proportions of salient regions.

Table 1: Proportion of salient $19x19$ pixel regions detected by GBVS at 30% threshold and selected by users, with resultant practical and theoretical entropy per click point. SR = salient regions. Bolded rows represent the images that were considered suitable.

Image	# SR	% Image salient	% selected SR	Theor. entropy	Prac. entropy
1	20	2%	100%	4.32	4.32
2	15	2%	100%	3.91	3.91
3	**146**	**24%**	**84.9%**	**7.19**	**6.95**
4	**242**	**45%**	**82.2%**	**7.92**	**7.64**
5	84	14%	88.1%	6.39	6.21
6	**236**	**42%**	**82.2%**	**7.88**	**7.60**
7	**176**	**32%**	**85.2%**	**7.46**	**7.23**
8	**153**	**23%**	**85.6%**	**7.26**	**7.03**
9	**242**	**45%**	**73.6%**	**7.92**	**7.48**
10	**270**	**45%**	**78.5%**	**8.08**	**7.73**
11	**279**	**47%**	**71.7%**	**8.12**	**7.64**
12	116	14%	83.6%	6.86	6.60
13	**187**	**28%**	**85.6%**	**7.55**	**7.32**
14	98	13%	74.5%	6.61	6.19
15	**197**	**30%**	**75.6%**	**7.62**	**7.22**
Avg.	164.1	27.1%	83.4%	7.01	6.74

We then used Equation 1 as defined in [4] to compute the entropy per click point of the user-created passwords, where b represents the number of salient $19x19$ regions, and k is the length of the password. In this case, the entropy was computed for each actual click point in turn, so $k = 1$.

$$H(I) = Log_2(b^k) \quad (1)$$

Our hypothesis is that the larger the proportion of the image that is considered salient, the larger the number of salient regions users may choose, and thus the higher the practical password space for that image. We calculated entropy of the *theoretical* password space using the total number of salient regions in an image, and the entropy of the *practical* password space that considers the number of salient regions that were selected by users in the same image. The results of computing the entropy are shown in Table 1; the results show that some images such as images 10, 11, 4, and 6 have high entropy per click point because of the large proportion of the images that are covered by saliency maps, thereby providing a large number of $19x19$ regions that can contain click points. For instance, Image 2 (see Figure 2a) has a small salient proportion, and also a lower theoretical and practical password space.

We examined whether there is a correlation between the detected number of salient regions and those regions selected as part of a password. We computed the Pearson correlation between the entropy of the two variables in Table 1. The result indicated that there was a strong, positive, statistically significant correlation between the entropy per click point of total salient regions and the entropy per click point of the selected salient regions ($r(15) = 1$, $\rho = 0.0$). This implies that users tend to choose different salient regions when there is a large proportion of salient regions that cover the image.

The high degree of correlation between the entropy of total and selected salient regions indicated an image with more salient regions has more prominent hotspots and a larger selection of appropriate graphical password click points. As a result, such images with highly salient regions can lead to more secure password selections, which provides support for our hypothesis in the previous paragraph.

We formalize this result into a decision model: if an image A has entropy value E_A and an image B has entropy value E_B, and $E_A > E_B$, then image A is more suitable than image B as a guiding image for graphical passwords. The result of this stage is that the higher the image saliency proportion for the guiding image, the higher the entropy of the resulting password. Therefore, discarding the images with the lowest proportion of saliency is a promising way to encourage users to create stronger passwords. We determined empirically that a total image saliency percentage threshold of 23% (see the % image salient column in Table 1) produced the best results in terms of determining whether or not an image was suitable for use with graphical passwords. We define 'best' here as having balance between image saliency proportion and the resulting theoretical and practical password entropy, as well as confirming visual determinations of saliency and thus suitability. This threshold depends on the input images, and therefore will change depending on the images chosen.

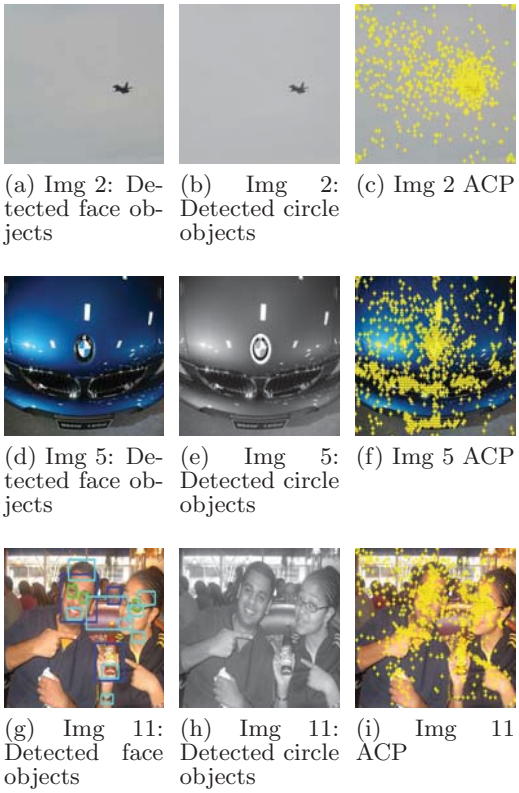
3.3 Stage 2: Detecting Image Objects

After removing the least suitable images using the method described in Stage 1, we further refined image suitability by relating the salient regions to actual objects in the image (which we call Regions of Interest, or RoI) that drew the participant's eye when creating their graphical passwords. Our rationale for exploring this relationship is that simple saliency does not determine whether an image region will be selected by a user, and thus there must be some other factor that encourages such a selection. We hypothesize that objects, which are likely to have meaning to a user and thus be perceived as more memorable, may be this factor in encouraging selection. As an example, salient regions may to be those that have changes in contrast, such as a "salt and pepper" image of random white and black pixel regions. However, a simple change in contrast does not necessarily have meaning to a user and therefore may not be memorable in terms of point selection for a graphical password. To relate saliency to objects, we segmented the images into $19x19$ pixel squares and detected objects within each square (see Figure 2 for examples). If a square contained at least one detected object, we consider it an RoI.

We then computed the likelihood that a user will select the detected region based on how frequently it was selected and use this likelihood to compute the entropy of the resulting passwords. We evaluate our model using real graphical passwords chosen by users in the study by Zhao *et al.* [37].

We based our RoI detection method on the way users create graphical passwords in Windows 8^{TM} [23]. Here, users choose from three actions as part of their password: a single tap, a line, or a circle. Given these actions, we expect that users will choose regions on the guiding image that correspond to the available actions. For instance, objects such as faces, eyes, noses, and mouths may be the target of taps, or the user may draw a line from one of these objects to another. Moreover, a detected circle in the guiding image could encourage users to trace a circle over the object, or to tap inside the circle. These objects have been shown to be attractive to users when creating graphical passwords [37].

Figure 2: Output samples of the detected objects and circles and the ACP chosen by users in Zhao *et al.*'s study [37]. Background images originally from [14].



(a) Img 2: Detected face objects (b) Img 2: Detected circle objects (c) Img 2 ACP

(d) Img 5: Detected face objects (e) Img 5: Detected circle objects (f) Img 5 ACP

(g) Img 11: Detected face objects (h) Img 11: Detected circle objects (i) Img 11 ACP

3.3.1 *Face object detection*

We used the Viola-Jones algorithm [34] with Haar cascades [1] to detect six different object classes: face, eye, nose, mouth, left ear and right ear. Any object that is detected within the boundary of the face is considered part of the face. However, any object that is detected as one of the face parts (eye, nose, etc.), but is located outside the face boundary is considered a false positive, so we called it a "Generic" object. Figures 3a, 3d, and 3g show examples of detected objects for a sample of the training images.

3.3.2 *Circle object detection*

We used the Java implementation of the Hough Transform Circle detection algorithm [3, 25] to detect the location of the circle's center points and radii lengths. Figures 3b, 3e, and 3h show examples of circles detected by this method.

3.3.3 *Image corners*

As can be seen in Figures 3c, 3f, and 3i, many users clicked on the image corners themselves (top left, top right, bottom left and bottom right, as opposed to corners that appear in the subject of the image, such as a building) even though there may not have been a detected region of interest. We hypothesize that this is because these points are memorable

Table 2: Proportion of actual and detected click points and their related entropy per click point in bits. We show the results for all 15 original images for comparison purposes, although those considered unsuitable in Stage 1 (unbolded) would normally not be shown to users when creating passwords.

Image	# ACP	# Detected CP	Proportion Detected	Entropy per CP
1	1995	756	38%	0.42
2	2007	135	7%	0.30
3	**1986**	**1471**	**74%**	**1.22**
4	**1995**	**1585**	**79%**	**5.88**
5	2001	749	37%	0.74
6	**2028**	**1245**	**61%**	**5.56**
7	**2019**	**203**	**10%**	**0.59**
8	**2013**	**1250**	**62%**	**2.82**
9	**2016**	**1639**	**81%**	**3.99**
10	**2013**	**881**	**44%**	**1.64**
11	**2013**	**1262**	**63%**	**2.24**
12	2007	1325	66%	3.89
13	**2010**	**1344**	**67%**	**3.12**
14	2010	1112	55%	1.28
15	**2004**	**906**	**45%**	**1.36**
Average	2007.8	1057.5	53%	2.34

on all images, regardless of the image's guiding features. Thus, we consider these $19x19$ pixel regions in the image corners as RoI.

3.3.4 Segmentation and Labeling

Each $19x19$ pixel RoI was labeled with a detected object type based on its location. For example, an RoI is labeled as an eye object if the region falls within the boundary of a detected eye object. If the RoI covers two different types of detected object because of overlap, the RoI is labeled as containing the object that has shortest path from the center point to the RoI's center point. If an RoI has no detected objects within it, it is labeled with NONE.

After labeling each $19x19$ pixel region, we relate this to the ACP that comprise real passwords as taken from the training data. For each ACP, it belongs to that object if the point falls within the object's boundary. Furthermore, each selected point may belong to at most one object. If the point falls into more than one object's boundary because of overlapping objects, the selected point belongs to the object that has shortest path from the point to the object's center point.

3.4 Relating Detected Objects to ACP

Table 2 shows that the detected RoI predicted an average of 53% of ACP. The lowest percentages were in images 2, 7, 5, and 1 and we believe that this is due to the image's content. For example, image 2 (see Figure 3a) is simple (a small jet against the otherwise unremarkable sky), and image 5 (see Figure 3d) has a small car logo against an equally unremarkable car hood. Since these images have few hotspots, users may be forced choose regions that are not detected as RoI. The highest number of ACP appearing in detected objects was in image 9 with 81%. The faces of two people cover the majority of image, which provides many detected objects that may also encourage clicks.

None of the images had all ACP related to a detected object. This may be due to a semantic gap between what is detected on an image and what the user perceives as important or memorable. It is unlikely that this intuition can be automated unless input from users is applied. Another explanation may be because ACP appear in areas where there are no objects to detect, but where the image bounds provide guidance. For instance, we have noticed that a proportion of ACP fall in the image corners for all 15 images, even though there are often no objects to draw the user's attention. We have (somewhat artificially) added these as "detected" objects because they are easily located in the image simply by locating the image's boundaries. However, many of the images in Figure 2 show ACP that are not over a detectable object or otherwise easily located via image boundaries. We have no knowledge of why the user chose such points as they are unlikely to be memorable in the future.

The semantic gap between clicks and object detection is a significant component of this work. The object detection algorithms are only able to predict what might be a face, circle, etc., but not whether the user's attention is actually drawn to it. Images may contain features that make sense for users, perhaps due to other memorable links, but the user's semantic representation was not replicable by the object detection algorithms. For example, the image's corners are memorable due to the image boundaries, but may not contain detectable RoIs.

3.5 Region Selection Likelihood

As shown in Tables 2 and 3, the ACP were frequently on objects we detected in the fifteen images. Intuition leads us to use the click frequency of an RoI as an indication of the likelihood of a user clicking on that RoI in the future. However, simply calculating a likelihood that an actual click point appears in a given RoI without considering the RoI's actual content may lead to inaccuracies. For example, there is high likelihood that faces will be chosen in images that contain faces (e.g., Image 11), but this behavior is not expected for images with no faces (e.g., Image 5). As a result, calculating the percentages of RoIs with no consideration of the content of the images (as in Table 2) will lead to unrealistic assumptions. For instance, 33% of click points were on a detected face object in Image 3 (see Table 3), but only 3% of ACP were on a detected face in Image 7. Thus, the average of 8.3% of ACP is not appropriate for all images since differences in the other RoIs or the proportion of the chosen face to other objects may have an effect on whether or not the user clicks on any one object detected in the image. In order to calculate a more accurate likelihood that a user may click on a particular object, we considered two issues: the size of detected objects and the variety of object types detected. For instance, Image 6 has only circle objects detected, and Image 8 has face and circle objects (if we ignored image corners for both images). Table 3 shows that the likelihood of clicking on a circle object for Image 6 is 55%, but for Image 8 is only 4%. The reason could be that the size of circle object in Image 6 is larger than in Image 8. Not only that, but also Image 8 has different other objects such as face that may be a target for users' taps, while Image 6 has no other objects that may attract taps. As a result, the size of detected objects and the content of images should be considered for accuracy.

3.5.1 Image Categorization

In order to consider similar images (i.e., those with similar RoI content), we categorized the fifteen images based on their content. We used the type of detected objects to categorize the images where F represents a detected face, G a detected generic object and C for a detected circle. Combi-

Table 3: Percentages of ACP related by the different types of objects per image. Blanks imply 0%.

Img	Face	Eye	Nose	Mouth	Generic	Circle	Corners
1					31%		6%
2							7%
3	33%		11%	24%			6%
4					9%	65%	5%
5						32%	6%
6						55%	6%
7	3%				2%		5%
8	21%		14%	7%	1%	4%	6%
9	12%	8%	22%	8%		26%	5%
10	32%			2%	2%		7%
11	24%	7%		5%	21%		6%
12					33%	27%	6%
13					13%	49%	5%
14					49%		6%
15					39%		6%
Average	8.3%	1.0%	3.1%	3.1%	13.9%	17.2%	5.9%

Table 4: The number of regions covered by detected objects aggregated by the object type and category.

Category	Face	Eye	Nose	Mouth	Generic	Circle	TL	TR	BL	BR
F	4	0	4	11	0	0	1	1	1	1
FG	35	7	0	16	47	0	3	3	3	3
FGC	11	10	11	27	43	264	3	3	3	3
G	0	0	0	0	58	0	3	3	3	3
GC	0	0	0	0	30	161	2	2	2	2
C	0	0	0	0	0	82	2	2	2	2
NONE	0	0	0	0	0	0	1	1	1	1

nations of these objects are represented as concatenations of these three; for instance, FGC represents images that have detected faces (F), generic objects (G) and circles (C). The fifteen images are categorized into eight possible categories, see Table 4.

No images contained only face and circle objects, so we ignored this category and use only seven categories. The number of regions and the ACP on the fifteen images are shown in Tables 4 and 5, respectively. We used them to compute the percentage of ACP for each RoI by computing the total number of ACP that fall into each RoI for each category (see Table 5). Since the image corners have a different number of ACP based on their location, we consider them separately: top-left (TL), top-right (TR), bottom-left (BL), and bottom-right (BR).

The percentage for each region is calculated by dividing the total number of ACP on each category of images aggregated by the object type as seen in Table 5 by the total number of regions in the same category as indicated in Table 4. The result is the average of ACP per region which is divided by the average number of ACP that have been made on each image. Table 6 shows the percentage of the average of ACP per region based on the object type and category of the image. The lower the percentage for a given category in Table 6, the less likely it is that a user will select that location. For instance, in the FG row in Table 6, which represents images containing faces and generic objects, users are unlikely to select a Nose object (0% likelihood), but are more likely to select the top left corner (6%).

3.6 Password Entropy

As with the saliency calculations in Section 3.2, we used entropy as a strength measure for the graphical passwords created by the users in Zhao *et al.*'s study. We assigned a likelihood of selection to each $19x19$ pixel region in each of the fifteen images based on the method described in Section 3.5. We used this likelihood as input to calculating the practical password space using entropy using the Shannon equation [4], which is measured in bits.

Table 5: The number of ACP that fall in the detected object aggregated by the object type and category

Category	Face	Eye	Nose	Mouth	Generic	Circle	TL	TR	BL	BR
F	6	0	59	225	0	0	110	2	0	0
FG	302	165	0	239	272	0	357	3	2	3
FGC	47	55	276	759	111	1937	326	1	3	3
G	0	0	0	0	596	0	343	6	6	7
GC	0	0	0	0	156	1528	223	0	0	1
C	0	0	0	0	0	1788	215	4	4	1
NONE	0	0	0	0	0	0	112	4	5	6

Table 6: The likelihood of each region aggregated by the object type and category. The empty cells have zero likelihood.

Category	Face	Eye	Nose	Mouth	Generic	Circle	TL	TR	BL	BR
F	<1%		1%	1%			6%	<1%		
FG	<1%	1%		1%	<1%		6%			
FGC	<1%	<1%	1%	1%	<1%	<1%	5%			
G					1%		6%	<1%	<1%	<1%
GC					<1%	1%	6%			
C						1%	5%	<1%	<1%	
NONE							6%	<1%	<1%	<1%

Table 2 shows the entropy calculations per click point for the 15 images in the dataset. Higher entropy implies a higher practical password space, which means that there are more possible passwords that must be guessed during an attack. The highest entropy (5.88 bits per click) was for Image 4, which means that this image is the most complex image (in this dataset) since it has more regions that are less frequently selected compared to more frequently selected regions. Images 1 and 2, both of which show a jet against the sky had the lowest entropy at 0.30 and 0.42 bits per click, respectively, which means they are the least suitable images because they have an insufficient number of RoIs to allow for selected points to be spread across the image. Therefore, the lack of RoIs in an image may lead users to create (likely) memorable but also predictable (e.g., weak) passwords, and the existence of many and varied RoIs may lead users to create stronger but (likely) less memorable passwords.

We computed the coefficient of multiple correlation to measure how the entropy of practical password space can be predicted using a linear regression equation of a set of variables that represent the number and type of RoIs. The statistically significant result indicated that the number and type of regions predicted the entropy of practical password space, $F(6,8) = 7.049; \rho < 0.05; r^2 = 0.841$. The value of the multiple correlation coefficient $R = 0.92$ means that the model has a good level of prediction. Equation 2 shows the model, where F, E, N, M, G, C are the number of detected RoI in an image for face, eye, nose, mouth, generic, and circle, respectively. Note that the weights in Equation 2 were experimentally determined during the regression calculation; they will vary depending on the input images used. The model simplifies computing the predicted entropy of practical password space for an image because it depends on the number and type of detected RoIs.

$$H(I)_{pred} = W_1 - (W_2 * F) + (W_3 * E) - (W_4 * N) + (W_5 * M) + (W_6 * G) + (W_7 * C) \quad (2)$$

As a comparison point, the per-character entropy for a nu-

meric PIN is 3.32 bits per digit, for a case-sensitive password with letters only is 4.7 bits per character, and for a textual password made up of case insensitive letters (52 choices), digits (10 choices) and special characters (10 choices) for a total of 72 possible characters to choose from is 6.17 bits per character. Therefore, the per-bit entropy for the best images in our study (Images 4 and 6) were greater than that of numeric PINs and case-sensitive passwords, but smaller than that of stronger passwords that contained more characters. This shows the possibility for strong graphical passwords with these images, particularly if we consider increasingly complex images with more possible objects that could be selected.

4. USER STUDY

While the results from using the dataset from Zhao *et al.* [37] provided base information for creating a suitability measurement for graphical password guiding images, there were still several outstanding questions regarding usability and memorability that could not be answered using Zhao *et al.*'s dataset. Therefore, we undertook a user study (approved by our institution's IRB prior to its start) to gather user data to answer questions the following questions:

1. Do participants tend to choose suitable images more or less frequently than unsuitable images, where "suitable" and "unsuitable" are determined by the image saliency Stage 1 procedure described in Section 3.2?
2. Do participants select $19x19$ regions evenly, or are there certain regions that are selected more frequently than others?
3. Do participants find graphical passwords easy or hard to use?
4. Are passwords created using a more suitable guiding image more memorable than those created with a less suitable image?

The answer to question 2 above will have an effect on the determination of practical password space, as defined previously in this paper. Theoretical password space assumes that, of the regions actually chosen, that all of these are equally likely to be chosen. Therefore, we wished to further bound the theoretical password space estimates given in Table 2 by providing information on how frequently each chosen space was actually used in a password, thereby giving an idea of whether the original theoretical password space also overestimates the difficulty in cracking a given password.

4.1 Study Design

We developed an Android application for creating graphical passwords and ran a two-session user study designed to answer our research questions. We recruited a total of 33 participants (6 female and 27 male, mean age = 30.18 years) via email and personal invitation. Participants were not required to have experience with graphical passwords, although we required them to have experience with touch-based mobile devices. The first session, in which participants enrolled with their chosen password and answered demographic and post-enrollment questionnaires, lasted around 30 minutes; the second session, which was to recall their password after one week, lasted less than 5 minutes. The sessions took place in a quiet environment with participants seated in a chair throughout.

4.2 Procedure & Equipment

In the first session, participants enrolled with their chosen password. They first chose a guiding image from the set of 15 images used in the Zhao *et al.* study [37]. Each chosen image was reviewed using the image saliency procedure in described in Section 3.2; if the image was considered unsuitable, the participant was prompted to select a different image. This process was used to find out how frequently participants chose the unsuitable images from the original set of 15 images; it is our intention that the saliency procedure would be used to filter out images that would not be shown to participants, we were not interested in studying passwords created with these unsuitable images. Once the participant had chosen a suitable guiding image, they created a graphical password consisting of three points. Each of the three points could be a tap, a line or a circle anywhere on the guiding image and they were allowed to start again if, for some reason, they did not like their original password. After creating the password participants played a memory rotations game to clear their short-term visual memory, and then were asked to recall their password. We then asked a few questions about their experience in a semi-structured interview. During the second session, which was approximately one week after the first session, participants were once again asked to recall their password, but did not play the mental rotations game nor have an interview.

Participants used an LG Nexus 7 tablet when participating in both sessions. We chose to use a tablet rather than a smartphone due to the larger screen; as an initial study we did not want to conflate the usability of the device itself with the creation of the password itself. The tablet had loaded a bespoke app designed for the study. Its purpose was to first provide an interface for the participant to create and recall their graphical password, as well as gathering data such as the participant's demographic information, the guiding images chosen, and the locations and types of actions chosen by the user to make up their graphical password. We used this app to gather timing information for the various stages of the study including password creation and recall.

4.3 User Study Results

We now present the results of the user study in order to answer the questions that we raised in this section. Our results are categorized in terms of image selection, usability, and password memorability.

4.3.1 Image Selection

To determine the answer to the first research question regarding whether participants chose suitable or unsuitable images more frequently, we examined the choices of images the participants selected as guiding images for their passwords. Out of the 15 images initially provided, participants selected only seven images, as shown in Figure 3. Figure 4 shows the total selections by the 33 participants. Note that the totals add up to more than 33 selections because we are also showing the initial unsuitable choices (blue bars, images 1, 2, 5, 12, 14, 15). 11 out of 33 participants selected an unsuitable image and were required to select another image before progressing. Given that the other 22 participants chose a suitable image initially, this shows that participants

in our study were more likely to choose a visually more interesting image even when simpler images were available.

Figure 3: Images selected by the participants in the user study. Background images originally from [14]

(a) Img 4 (b) Img 6 (c) Img 7 (d) Img 8

(e) Img 9 (f) Img 11 (g) Img 13

After the participant chose a suitable image, we logged how frequently each of the seven images in Figure 3 were chosen. Comparing the most frequently chosen images to the results of entropy calculations for each image in Table 2, we see that the most frequently chosen images were also those with the highest per-click entropy (Images 4 and 6). This means that most of the participants in our study chose the most suitable images, which is encouraging since they are also likely to make the strongest passwords.

Now, we examine only those images that were considered suitable by the Stage 1 image saliency measurement in addressing usability and memorability.

4.3.2 *Usability and Memorability*

To measure usability, we recorded password creation time and password recall time. As can be seen in Figure 5, the highest average creation time was around 28 seconds and the lowest around 11 seconds. The login and recall times are lower for most images, with the recall time after one week quite close to the initial recall (login time) immediately after creating the password. It is acceptable to have a longer creation time since creation happens once; it is therefore a positive result to see that login and recall times are somewhat lower, with the one-week recall time similar to the immediate recall (login) time in most cases. This shows that the passwords created were relatively easy to enter, which provides evidence of the ease of password use in our third question in Section 4. We note, however, that these times are likely higher than required to enter a textual password, although this may be due to lack of familiarity and repetition with our graphical password method. Interestingly, Image 4, the second most suitable image, had a high creation, login and recall time compared to other images. However, the most suitable image, Image 6, had comparatively low creation, login and recall times. This rather confusing result is likely due to the specific image contents.

Figure 4: The totals of participants who selected each image in the user study. Blue bars represent unsuitable choices per Stage 1.

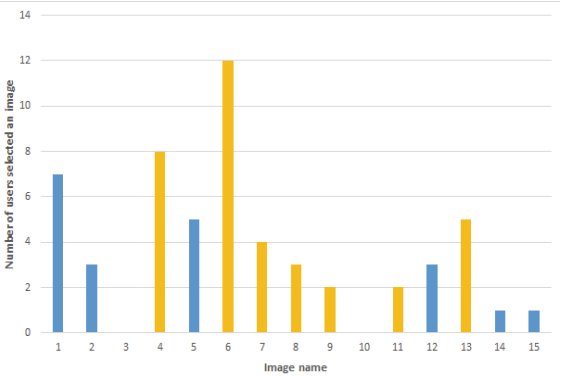
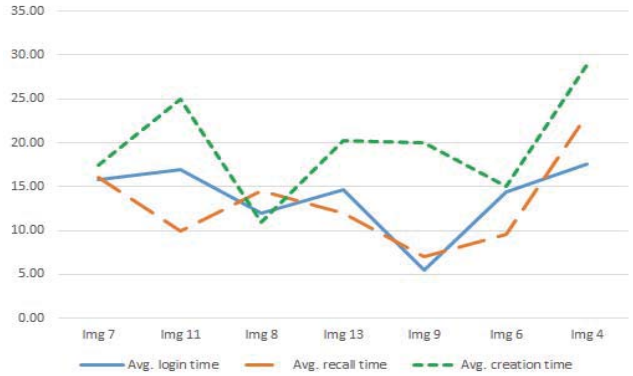


Table 7: The Kruskal-Wallis test results for login, recall, login-retry, and recall-retry. df = 6, N = 33, and critical value = 12.5916, alpha = 0.05

	H	Significant difference?
Login	-23.3805	No
Recall	-24.8478	No
Login-Retry	-98.7848	No
Recall-Retry	-100.166	No

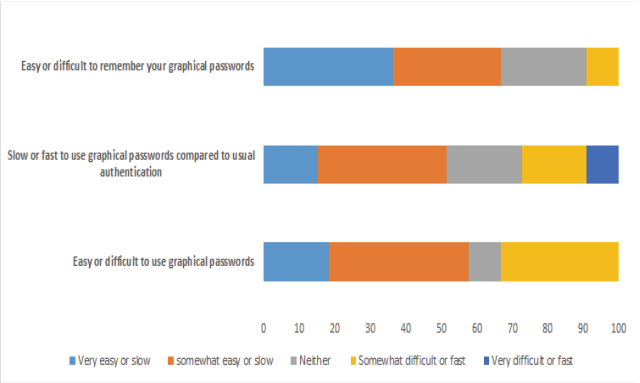
Figure 5: The average time of creating (small dashed line), logging (solid line) in and recalling (large dashed line) graphical passwords for participant-selected. Images are ordered from least suitable to more suitable.



The time that participants took to login or recall the passwords and the number of times that participants needed to successfully enter their password were measured and assessed with the Kruskal-Wallis test in order to see if there is any difference between the images selected. Table 7 indicated that there was no statistically significant difference in terms of login and recall times, nor the number of times needed to successfully authenticate during initial login (login-retry), and during the one week recall (recall-retry) among the seven images. As a result, using the most suitable images did not affect login or recall times, and thus does not have a negative effect on graphical password usability.

Another important feature of usability is user satisfaction with the proposed scheme. To measure this, we conducted an interview with participants after the initial creation and recall phases were complete. Figure 6 shows the responses to the questions asked as they relate to user satisfaction. The

Figure 6: The responses of participants related to usability of graphical passwords



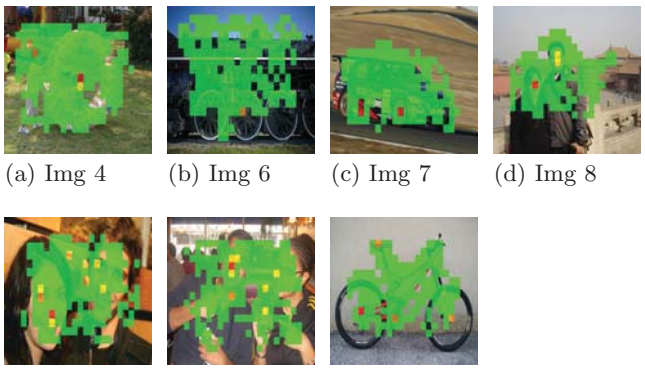
majority of participants ($\tilde{6}8\%$) found it easy or very easy to remember their password, while the participants were split quite evenly on whether they found graphical passwords slow to use when compared to their regular authentication method. A similar number, slightly more than half, of participants found graphical passwords easy to use. These results are somewhat non-deterministic: while participants found graphical passwords easy to remember, they were also considered no faster nor easier than traditional methods, which means that they are about as usable as traditional methods. While this result shows positive possibilities for memorability (when combined with the fact that most participants needed only one try to recall their password after one week), it shows little improvement in overall usability. Therefore, we have shown support for our fourth research question regarding memorability since less suitable images showed similar creation and recall times to more suitable images. However, we also consider this result heartening since we do not believe the participants considered the graphical password less usable than traditional methods, and they were overall less familiar with our method.

4.3.3 User Choice

Finally, to answer question 2 from Section 4 regarding user choice of RoIs, we created a heatmap (see Figure 7) for user-selected images to show how frequently each RoI was chosen. The colors range from green (lightly shaded) for the least frequently selected regions to yellow and red (darker shading) as the square is more frequently chosen. If a RoI was not chosen, or if the region was not considered an RoI since it did not contain a detected object, the corresponding square is not colored or shaded. Referring to the heatmap of Image 4 (Figure 8a), which is one of the two most suitable images, we see that the center of the circular object was frequently selected (red/darkly shaded square) with a yellow square just below it. This could be due to participants selecting the center but with less accuracy. Similarly, we see in the heatmap for Image 8 (Figure 8d) that the two faces in the image were frequently selected, which corresponds with previous results that showed that faces were attractive regions [37].

Overall, the heatmaps in Figure 7 show that users tend to select RoIs relatively evenly, as shown by the preponderance of green/lightly shaded RoIs on each image. This shows that, while hotspots do exist in each image, they are not necessarily preferentially chosen by users. This supports our intuition from Section 3.2.1 that more RoIs may encourage users to select regions that are not necessarily hotspots, but that are still visually interesting and potentially memorable. We chose not to consider non-RoI squares in this assessment as we believe that if there is no detected RoI that these selections will not be memorable. We intend to test this assumption in future work.

Figure 7: Sample of heatmaps for images that represents selected areas. Frequencies ranged from red (dark) for most frequently selected to green (light) for least frequently. Background images originally from [14].



(a) Img 4 (b) Img 6 (c) Img 7 (d) Img 8

(e) Img 9 (f) Img 11 (g) Img 13

5. CONCLUSION

Graphical passwords have become a plausible authentication approach because of their usability in terms of login error rate [6]. However, developing a method to evaluate the predictability of graphical passwords is essential to maintain the strength at least equal to that of textual passwords. In this work, we add to the literature on graphical password strength by proposing a measurement for guiding images that is based on overall image saliency and contents. Our measurement can be used to guide suitable image selection for graphical passwords before the user selects their first click point and without requiring additional user effort, which is our main contribution. Our main message is that graphical passwords are well suited to the largely touch-based input common on modern mobile devices, are more memorable than textual passwords, and therefore are likely to have more scope for future research compared to textual passwords. Research to discover improvements in graphical password security, memorability and usability may create systems that are better suited to keyboardless or hard to type on devices, where traditional textual passwords are unlikely to improve.

We evaluated our saliency measurement model based on existing graphical password click points, and we found that the more salient regions on an image, the higher the entropy of click points, and thus the higher the theoretical and practical password space. We then took the most suitable images from this evaluation and determined that the resultant passwords will be stronger if a guiding image with more RoI is chosen. In future work, we would like to broaden the participant pool to include more diverse users in order to

generalize the results presented here, as well as asking more in-depth questions regarding user reasoning behind image choices. We also intend to further support our hypotheses regarding image saliency and regions of interest with images other than those used by Zhao *et al.*

Another important contribution of this work is a challenge to the belief that hotspots are a detriment to graphical password guiding images. We show that a more complex image (i.e., one with more RoIs) can provide more possible click points that users actually select, thus potentially encouraging the user to create a less guessable password, or at least one that is more computationally intensive and time consuming for an attacker. This has the effect of spreading potential user click points, and thus potential hotspots, over an image, which may make dictionary attacks that need to consider all hotspots less feasible.

Future extensions to this include considering other object detection algorithms and color features that may provide additional information about the regions that are likely to be selected by users. Additional future work will study whether images with more RoI encourage users to choose more secure and memorable graphical passwords. This will be explored with the addition of a graphical password strength meter based on the strength metric proposed in this work.

6. REFERENCES

[1] Alerimondo.no-ip.org. Haar Cascades. http://alereimondo.no-ip.org/OpenCV/34, 2014.

[2] J. M. Allen, L. A. McFarlin, and T. Green. An In-Depth Look into the Text Entry User Experience on the iPhone. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 52(5), pages 508 – 512, 2008.

[3] D. H. Ballard. Generalizing the Hough Transform to Detect Arbitrary Shapes. *Pattern Recognition*, 13(2):111 – 122, 1981.

[4] W. E. Burr, D. F. Dodson, E. M. Newton, R. A. Perlner, W. T. Polk, S. Gupta, and E. A. Nabbus. Electronic Authentication Guideline. Technical Report NIST Special Publication 800-63-2, NIST, 2013.

[5] S. Chiasson, A. Forget, R. Biddle, and P. van Oorschot. Influencing Users Towards Better Passwords: Persuasive Cued Click-Points. In *Proceedings of the 22nd British HCI Conference*, volume 1, pages 121 – 130, 2008.

[6] S. Chiasson, A. Forget, E. Stobert, P. C. van Oorschot, and R. Biddle. Multiple Password Interference in Text Passwords and Click-Based Graphical Passwords. In *Proceedings of the 16th ACM Conference on Computer and Communications Security (CCS)*, pages 500 – 511, 2009.

[7] S. Chiasson, E. Stobert, A. Forget, R. Biddle, and P. van Oorschot. Persuasive Cued Click-Points: Design, Implementation and Evaluation of a Knowledge-Based Authentication Mechanism. *IEEE Transactions on Dependable and Secure Computing*, 9(2):222 – 235, 2012.

[8] S. Chiasson, P. C. van Oorschot, and R. Biddle. Graphical Password Authentication Using Cued Click Points. In *Proceedings of the 2007 European Symposium on Research in Computer Security*, volume 4734/2007 of *Lecture Notes in Computer Science*, pages 359 – 374. Springer Berlin / Heidelberg, 2007.

[9] D. Davis, F. Monrose, and M. K. Reiter. On User Choice in Graphical Password Schemes. In *Proceedings of the 13th Conference on USENIX Security Symposium*, volume 13, pages 151–164. USENIX, 2004.

[10] X. de Carnede Carnavalet and M. Mannan. From Very Weak to Very Strong: Analyzing Password-Strength Meters. In *Proceedings of Network and Distributed Systems Security Symposium (NDSS)*, 2014.

[11] A. E. Dirik, N. Memon, and J.-C. Birget. Modeling User Choice in the PassPoints Graphical Password Scheme. In *Proceedings of the 3rd Symposium on Usable Privacy and Security (SOUPS)*, pages 20 – 28, 2007.

[12] P. Dunphy and J. Yan. Do Background Images Improve Draw-a-Secret Graphical Passwords? In *Proceedings of the 14th ACM Conference on Computer and Communications Security (CCS)*, pages 36–47, 2007.

[13] R. English and R. Poet. Towards a Metric for Recognition-Based Graphical Passwords. In *Proceedings of the 5th International Conference on Network and System Security (NSS 2011)*, page to appear., 2011.

[14] M. Everingham, L. V. Gool, C. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, 88(2):303 – 338, 2010.

[15] W. B. Glisson, T. Storer, G. Mayall, I. Moug, and G. Grispos. Electronic Retention: What Does Your Mobile Phone Reveal About You? *International Journal of Information Security*, 10(6):337 – 349, 2011.

[16] K. Gołofit. Click Passwords Under Investigation. In *Proceedings of 12th European Symposium on Research in Computer Security (ESORICS)*, volume 4734 of *Lecture Notes in Computer Science*, pages 343 – 358. 2007.

[17] K. K. Greene, M. A. Gallagher, B. C. Stanton, and P. Y. Lee. I Can't Type That! P@$$w0rd Entry on Mobile Devices. In *Proceedings of HCI: Human Aspects of Information Security, Privacy and Trust*, volume 8533 of *Lecture Notes in Computer Science*, pages 160 – 171, 2014.

[18] J. Harel. A Saliency Implementation in MATLAB. http://www.klab.caltech.edu/~harel/share/gbvs.php, July 2012.

[19] J. Harel, C. Koch, and P. Perona. Graph-Based Visual Saliency. In *Proceedings of Neural Information Processing Systems (NIPS)*, 2006.

[20] L. Itti, C. Koch, and E. Niebur. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254 – 1259, 1998.

[21] L. M. Mayron. A Comparison of Biologically-inspired Methods for Unsupervised Salient Object Detection. In *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, pages 1 – 4, 2013.

[22] L. M. Mayron and M. N. AlShehri. Evaluating the Use of Models of Visual Attention to Predict Graphical Passwords. In *The 52nd Annual ACM Southeast Conference*, 2014.

[23] Microsoft. Personalize your PC.

http://windows.microsoft.com/en-us/windows-8/personalize-pc-tutorial, 2014. Accessed: 02-17-2014.

[24] R. Morris and K. Thompson. Password Security: A Case History. *Communications of the ACM*, 22(11):594 – 597, 1979.

[25] OpenCV Lover. Hough Circle Detection in JavaCV. http://opencvlover.blogspot.com/2012/07/hough-circle-in-javacv.html, 2014.

[26] T. Page. Usability of Text Input Interfaces in Smartphones. *Journal of Design Research*, 11(1):39 – 56, 2013.

[27] A. Salehi-Abari, J. Thorpe, and P. van Oorschot. On Purely Automated Attacks and Click-Based Graphical Passwords. In *Proceedings of Annual Computer Security Applications Conference*, pages 111 – 120, 2008.

[28] F. Schaub, R. Deyhle, and M. Weber. Password Entry Usability and Shoulder Surfing Susceptibility on Different Smartphone Platforms. In *Proceedings of 11th International Conference on Mobile and Ubiquitious Multimedia*, pages 13:1 – 13:10, 2012.

[29] F. Schaub, M. Walch, B. Konings, and M. Weber. Exploring the Design Space of Graphical Passwords on Smartphones. In *Proceedings of 9th Symposium on Usable Privacy and Security (SOUPS)*, page Article 11, 2013.

[30] E. Stobert, S. Chiasson, and R. Biddle. User-Choice Patterns in PassTiles Graphical Passwords. In *Annual Computer Security Applications Conference (ACSAC)*, 2011.

[31] C. Sun, Y. Wang, and J. Zheng. Dissecting Pattern Unlock: The Effect of Pattern Strength Meter on Pattern Selection. *Journal of Information Security and Applications*, 19(4–5):308 – 320, 2014.

[32] J. Thorpe and P. van Oorschot. Human-Seeded Attacks and Exploiting Hot-Spots in Graphical Passwords. In *Proceedings of 16th USENIX Security Symposium*, pages 103 – 118, 2007.

[33] P. C. van Oorschot, A. Salehi-Abari, and J. Thorpe. Purely Automated Attacks on PassPoints-style Graphical Passwordstyle Graphical Passwords. *IEEE Transactions on Information Forensics and Security*, 5(3):393 – 405, 2010.

[34] P. Viola and M. J. Jones. Robust Real-Time Face Detection. *International Journal of Computer Vision*, 57(2):137 – 154, 2004.

[35] Z. Wang, J. Jing, and L. Li. Time Evolving Graphical Password for Securing Mobile Devices. In *Proceedings of 8th ACM SIGSAC Symposium on Information, Computer and Communications Security*, pages 347 – 352, 2013.

[36] J. Yan, A. Blackwell, R. Anderson, and A. Grant. Password Memorability and Security: Empirical Results. *IEEE Security & Privacy*, 2(5):25–31, Sept. – Oct. 2004.

[37] Z. Zhao, G.-J. Ahn, J.-J. Seo, and H. Hu. On the Security of Picture Gesture Authentication. In *Proceedings of 22nd USENIX Conference on Security*, pages 383 – 398, 2013.