Leveraging Cognitive Factors in Securing WWW with CAPTCHA

Amalia Rusu and Rebecca Docimo
Fairfield University, Fairfield, CT, USA, arusu@fairfield.edu
Adrian Rusu
Rowan University, Glassboro, NJ, USA, rusu@rowan.edu

Abstract

Human Interactive Proofs systems using CAPTCHA help protect services on the World Wide Web (WWW) from widespread abuse by verifying that a human, not an automated program, is making a request. To authenticate a user as human, a test must be passable by virtually all humans, but not by computer programs. For a CAPTCHA to be useful online, it must be easy to interpret by humans. In this paper, we present a new method to combine handwritten CAPTCHAs with a random tree structure and random test questions to create a novel and more robust implementation that leverages unique features of human cognition, including the superior ability over machines in recognizing graphics and reading unconstrained handwriting text that has been transformed in precise ways. This combined CAPTCHA protects against advances in recognition systems to ensure it remains viable in the future without causing additional difficulties for humans.

We present motivation for our approach, algorithm development, and experimental results that support our CAPTCHA in protecting web services while providing important insights into human cognitive factors at play during human-computer interaction.

1. Introduction

Most users of the WWW today are familiar with CAPTCHAs, which are presented to them as machine-printed text or sound samples to be interpreted. CAPTCHAs are typically used to prevent automated programs from gaining access to various Web resources for the purpose of spamming or other illegitimate use. CAPTCHA is needed because of the sheer volume of spam crossing the Internet and the agility and tenacity of spammers [12].

Artificial Intelligence (AI) experts consider CAPTCHA a win-win situation and point out that CAPTCHAs are useful even when broken for the insights provided to the field of AI [29, 30]. While breaking CAPTCHAs can be useful for advancing the field of AI as well as Image Processing, Pattern Recognition, etc., the current usefulness of CAPTCHAs in protecting Web resources from widespread illegitimate use by automated programs must not be overlooked.

In this paper, we present the development of a new CAPTCHA-based Human Interactive Proofs (HIP) authentication system to protect services on the WWW. In our system, users are authenticated as humans to gain access to Web services by correctly interpreting a tree structure with handwriting samples transformed according to specific principles of cognitive psychology, explained in greater detail in the next sections (Figure 1). To correctly solve the challenge, the tree structure and handwriting samples must be segmented out and interpreted, a task that presents much difficulty for machines, while being trivial for humans. With this CAPTCHA, we further the work begun on handwritten CAPTCHAs [22, 23, 24]. As CAPTCHAs are currently a readily available, relatively low cost and easy to administer solution to protect Web resources, our goal is to provide a useful CAPTCHA to overcome the security and usability difficulties present in other CAPTCHAs [17, 21, 33, 34]. Such a CAPTCHA can also offer extremely valuable insights not only related to the parsing of handwriting. By using a tree structure with handwritten images, both of which must be parsed to pass the CAPTCHA test, we can also offer important insights for the fields of AI, Image Analysis, Graphics Recognition and others.

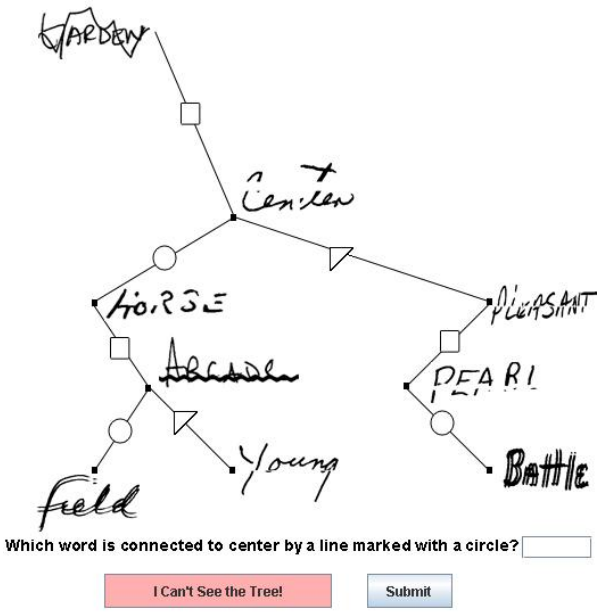


Figure 1. A tree-based handwritten CAPTCHA.

Tree drawings and handwriting are used in our CAPTCHA rather than the typical machine-printed text, both for the important advances to be gained in graphics and handwriting recognition fields if our CAPTCHA is broken, as well as for the security provided due to the extra challenges posed to machines. Human skill at interpreting basic drawings and handwriting, no matter the condition (i.e., rotated, occluded, or deformed) [20], is gained from an early age, while consistent machine recognition of graphics in general, and handwriting in particular, continues to be problematic mostly due to unconstrained writing style and segmentation, especially in the absence of a context [4, 7, 18, 26]. Moreover, when applying certain transformations to the handwriting and rendering it on a tree structure, the recognition drastically decreases. While humans are able to make use of certain aspects of perception and cognition to interpret transformed samples, this remains a difficult open problem for machines [9].

We begin by reviewing the concepts of both HIPs and CAPTCHAs. We then discuss advances in handwriting recognition and present human cognitive factors that relate to handwriting interpretation. In this context we introduce and discuss the Gestalt laws of perception and Geon theory related to human perception and reading skills to motivate the transformations we have applied to the images. The technical approach and methodology is then presented, as well as the findings of user studies and machine testing of our CAPTCHA to validate the usefulness of our system in protecting Web services. We conclude with important insights gained from our work and discuss possible future enhancements.

1.1 Overview of HIPs and CAPTCHA

The purpose of HIPs is to distinguish one class of users from another, most commonly humans from computer programs (also referred to as "machines") [1]. CAPTCHA is the test used by HIPs to distinguish a human user from a machine by presenting a challenge that can only be passed by the human. CAPTCHAs leverage AI factors and similarly to the tests of Alan Turing [28], they determine whether a user is human or machine. CAPTCHAs differ from Turing Tests, however, by making the computer the judge and administrator, though the computer itself should not be able to pass the test. In a CAPTCHA, if the individual completing the challenge presented passes correctly, they gain access to the service they are requesting. Otherwise, they are deemed to be an illegitimate program and are not allowed access. For a CAPTCHA to be useful, it must be easily passable by virtually all human users but not by machines [30]. If a CAPTCHA presents difficulty to machines, but also to humans, it has failed in its function [6, 21].

Primitive use of a commercial text-based riddle dates back to 1998 on the AltaVista search engine Web site (at altavista.com). Approximately two years later CAPTCHA was defined, along with the first commercial implementation by Carnegie Mellon University researchers. They set forth the basic properties of CAPTCHA: it must be automated so it can be administered by a computer program without human intervention, it must be public in that the test should not be unsolvable by machines simply because it is novel or the method or code is hidden, and it must be passable by virtually all humans but not by computer programs [30]. Many text-based visual CAPTCHAs have been created from Gimpy [10] to Baffletext [8] to reCAPTCHA [19]. An example text-based visual CAPTCHA is shown in Figure 2. Non-text visual CAPTCHAs have also been created, including those that leverage human cognitive abilities beyond word recognition, such as ARTiFACIAL [21] where users are asked to identify facial features.

Figure 2. Example of commercial CAPTCHA at msn.com.

1.2 Motivation for a Tree-based Handwritten CAPTCHA

While many CAPTCHAs have been created, more secure CAPTCHAs are needed to help secure the WWW against widespread abuse by programs posing as humans on the WWW. While a typical response to foil machine recognition to maintain the usefulness of CAPTCHAs is to make them harder for machines, care must be taken to ensure that human ease of use does not suffer. Accordingly, many current CAPTCHAs, both text and image based suffer from usability issues [6, 34]. Intrinsic security flaws of various CAPTCHAs have also been found and various text-based and image-based CAPTCHAs have been broken [7, 11, 33]. For example, Mori and Malik from the University of California at Berkeley demonstrate how they were able to break the EZ-Gimpy CAPTCHA with a 92% rate of success and Gimpy with a 33% rate of success [17].

It is the need for a more efficient CAPTCHA that is both usable for humans while secure against machines, along with the insights to be gained from persistent problems in computer recognition of handwritten text and graphics [5, 26] that motivates our approach. We combine a randomly generated tree structure with random test questions and mandatory interpretation of handwritten words transformed according to the Gestalt and Geon principles (Figure 1). This new challenge meets the criteria of being a CAPTCHA in that large quantities of human-like handwritten images can be automatically created via a synthetic handwriting tool [25] and then transformed and rendered on a tree structure. Our stringent adherence to the Gestalt and Geon principles related to human perception of objects, including letters and words, to create our transformations ensures that our new CAPTCHA is still easily solvable by humans while presenting challenges to computer programs. The tree-based CAPTCHA featuring deformed handwritten images (Figure 1) described in this paper addresses both security and usability aspects to create a viable alternative CAPTCHA to those in existence, while at the same time having the potential to add insights into the overlooked area of real time handwriting recognition and interpretation of complex multi-layer image based documents.

2. Technical Background

Challenge-response tests using visual recognition have been the most widely used type of CAPTCHA employed online to protect web services from abuse by automated programs. The purpose of this study and the approach developed is to expand the use of visual CAPTCHAs by inserting additional complexity for computers, while keeping the tests easy for humans to pass (again, if success rates for machines decrease but success rates for humans also decrease for a CAPTCHA, it is not viable [6]). In this context we discuss several factors that we have leveraged in our system.

2.1 Gestalt Principles and Geon Theory Factor

We have studied the Gestalt laws of perception and Geon theory and have used guiding principles of each to determine which very specific transformations can be applied to our handwriting samples to both assist human interpretation and present unique challenges to machine recognition. According to Gestalt principles, humans have a unique ability to make sense of pictures, even those that are incomplete or are marred in some way [15]. Humans are able to make sense of images presented to them by relying on their senses, past experience, which shapes how they view data currently, and what they are expecting to see. Humans are able to filter out irrelevant data such as noise or extra pieces in an image in order to interpret it. Gestalt principles are based on the fact that humans typically experience things that are outside of the range of simple perception. Humans tend to group information and interpret the whole rather than looking at individual pieces and then combining them. This is similar to the theory of holistic word recognition where the word is seen as an indivisible unit rather than as a series of individual parts which can be interpreted separately and then reassembled for recognition [16].

The Gestalt laws that aid human recognition of objects with transformations applied include proximity, similarity, symmetry, continuity, closure, familiarity and figure-ground as follows:

- Proximity: how objects are grouped together by distance from or location to each other.
- Similarity: how elements that are similar to each other tend to be viewed as part of a singular group.
- Symmetry: how objects are grouped into figures according to symmetry and meaning.
- Continuity: how objects are grouped according to flow of lines or alignment.

- Closure: how elements are grouped together if they tend to complete some pattern, allowing perception of objects that are visually absent.

- Familiarity: how elements are more likely to be interpreted as part of a group if they appear familiar to the viewer.

- Figure-ground distinction: how a scene is broken up into foreground (the object of interest) and background (the rest of the scene) which is what allows an object to be distinguished from its surroundings.

Other human cognitive factors at play in recognition are memory, internal metrics, familiarity of letters and letter orientation [22, 23, 24].

Human perception relies, in the end, on all of the Gestalt principles working together. In addition to using the Gestalt laws of perception to determine which transformations may be applied to CAPTCHAs to capitalize on machine recognition weaknesses and simultaneous human strengths, the Geon theory of pattern recognition is also useful to determine which core components must be present in a transformed image so that it is still interpretable by humans. Two key aspects of geons are edges and intersections. The importance of these has been tested on images where various parts were deleted [3]. Recognition for humans is easy if an object's geons can be recognized and edges and intersections are a critical part in recognition. We have made use of the Gestalt principles and Geon theory in development of our CAPTCHA through specific handwriting image transformations to ensure human legibility while foiling machines. Similar transformations can successfully be applied to the tree structure as well as any shape or object in general.

2.2 Handwriting Recognition Factor

Recognition of unconstrained handwriting, especially when it has certain transformations applied to it, continues to be a challenge for automatic recognition systems [26] while humans maintain a superior ability due to the Gestalt laws of perception [15] and Geon theory [2, 3]. Part of the problem for machines is that natural variability in handwriting exists at a level that does not exist in machine-printed text [26]. Our tests show that natural handwriting variability, as well as defects applied such as occlusions or fragmentations (Figure 4b), can currently be overcome by humans due to cognitive factors, but not by machines. Segmentation, or the ability for a machine to determine character and word boundaries, continues to be a problem [7, 18, 26]. Handwriting presents more segmentation issues for machines than machine-printed text making handwriting arguably superior to machine-printed text for use in a secure CAPTCHA.

While advances have been made in handwriting recognition and applications have found their place in certain contexts such as the US postal services [27], or bank check reading, these contexts are usually well known in the sense that a relatively small set of words is being used in a familiar and narrow context. Existing handwriting recognition approaches require a lexicon (as a dictionary or pre-determined list of words and expressions in a particular language used by a particular application), for high recognition accuracy [13, 14, 31, 32]. In our application to CAPTCHA, the use of words is infinite with no specific context, thus the required lexicon would have to be extremely large with consequently extremely poor accuracy by recognizers. Moreover, by applying very specific transformations that exploit the weaknesses of state-of-the-art recognition systems to our image samples on purpose, we add extra difficulty for machine recognition.

2.3 Graphics Recognition Factor

While advances have been made in the area of document image analysis, various open problems of interest remain. One key open issue is the lack of a general purpose cross-domain recognition tool. Most tools are very domain specific and require domain context [5] or a case-based approach [35] to interpret a particular graphic. For example, tools used to recognize domain-specific graphics such as electrical diagrams rely on primitives in the graphic that have intrinsic meaning. In musical scores, for example, the musical notes comprise a finite set of primitives that can be extracted and interpreted because they have a meaning apart from the whole graphic [5]. Once the primitives are interpreted, the graphic as a whole can be interpreted. We will discuss later how our tree drawing does not rely on any particular domain context and thus would be hard for machines to interpret. No parts of our tree have any intrinsic meaning and are always interpreted in the context of reading of deformed handwritten images.

3. Generation of Tree-based Handwritten CAPTCHA

The development of our CAPTCHA has focused on using transformed handwriting samples due to the aid

provided by cognitive principles that humans can make use of and the challenges presented to machines. We have increased difficulty for machines as well as the potential for insights in the area of graphics recognition by arranging our handwritten images into a tree structure. It should be noted that in our CAPTCHA the interpretation of the tree structure will always be combined with interpreting our handwritten images. There are two main parts to the creation of our combined CAPTCHA. First, images featuring synthetic handwritten are generated using a handwriting generator [25]. Gestalt and Geon-motivated transformations are then applied to the images to take advantage of human perception abilities and to create more difficulty for machines. Second, there is the random generation of a tree structure and tree elements and arrangement of the deformed handwritten images in the tree, making spatial recognition, which is a common task for humans, also part of the challenge. The overall architecture for our HIP system is shown in Figure 3.

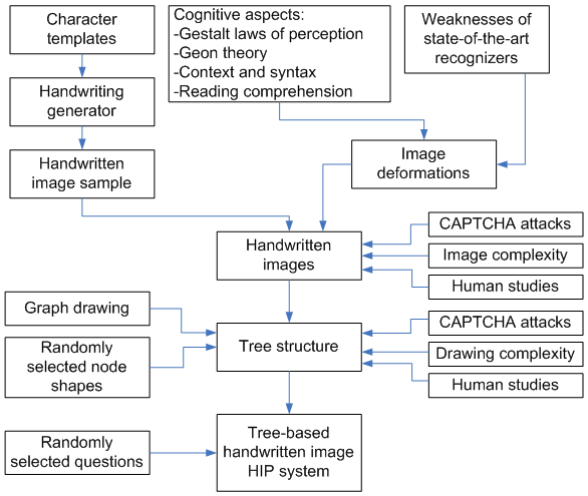
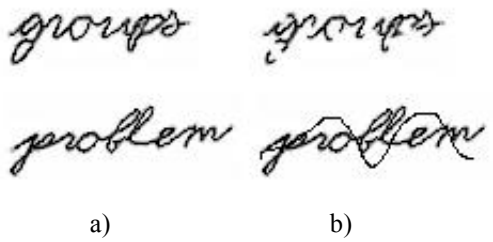


Figure 3. Overall architecture for tree-based handwritten CAPTCHA.

3.1 Transformed Handwritten Images

For testing our approach we have used handwritten word image samples. We have also developed a method to generate virtually infinite quantities of synthetic handwritten images based on real character templates [25] and to transform them on the fly. We note that a somewhat narrow set of words and their corresponding handwritten images was used for testing in order to provide machines with a lexicon to give them a fair chance at solving. Synthetic handwritten images that are generated and transformed on the fly for use in the CAPTCHA application are shown in Figure 4, before and after applying deformations that defeat state-of-the-art handwriting recognizers.



a) b)

Figure 4. Synthetically generated handwritten images: a) original; b) transformed.

To create our CAPTCHA, the number, type and severity of transformations are randomly chosen and applied, with some basic rules applied to ensure the images remain at once unreadable to machines and understandable by humans related to both Gestalt principles and Geon theory [2, 3, 15]. To the best of our knowledge, tying particular Gestalt principles to specific image transformations is novel and helps ensure that we maintain legibility for humans while foiling machine recognition.

Transformations that can be applied to handwritten images include horizontal or vertical overlaps with the principles guiding human interpretation proximity, symmetry, familiarity, continuity and figure-ground; occlusions such as by circles, rectangles or lines in the same color as the background pixels (Figure 4b, top) with the principles guiding human reconstruction closure, proximity, continuity and familiarity; occlusions by waves from left to right in the same color as the background with the principles guiding human reconstruction closure, proximity and continuity; occlusions using the same color pixels as the foreground with the principles guiding human reconstruction familiarity and figure-ground; adding extra strokes (Figure 4b, bottom) with the principles guiding human reconstruction familiarity and figure-ground; using empty letters, broken letters, rough contours, fragmentations, etc. with the principles guiding human reconstruction closure, proximity, continuity and figure-ground; splitting the image into parts and offsetting them from each other or splitting the image into parts and spreading the parts out (mosaic effect) with the principles guiding human reconstruction closure, proximity, continuity and symmetry; changing word orientation or stretching or compressing with the principles guiding human reconstruction memory, internal metrics, familiarity of

objects and orientation. The synthetic handwriting generation [25], as well as the transformations applied, ensures that infinitely many variations in samples can be produced to prevent machine learning.

3.2 Creation of a Tree Structure with Transformed Handwritten Images

Trees can be used for a wide range of tasks such as manipulating hierarchical data, making data easy to search, or as in our case, for presenting visual elements. Trees are drawn from the top down. A node is an element of the tree while a branch is a line that connects elements. We have used a binary tree in our CAPTCHA (Figure 1) to ensure the drawing does not become too large or unnecessarily complex for human readers. Each node in this tree has at most two branches. Our use of tree structures to add complexity to a handwritten CAPTCHA is motivated by several factors. First, we leverage currently superior human skills not only in reading carefully transformed handwritten images but in interpreting them inside a graphic with additional elements. Our CAPTCHA in all cases requires the interpretation of handwriting and the tree merely adds more complexity. Second, we take into account the various open issues in graphics recognition and in document analysis and recognition generally. Our use of tree drawings is further encouraged by early user studies, which indicate that the trees do not present additional complexity to humans beyond handwritten CAPTCHAs alone.

The tree generation algorithm for our CAPTCHA uses randomness whenever possible to create a random tree overlaid with transformed synthetic handwritten images. The program begins with generating a random number of nodes. Once the number of nodes has been determined, the algorithm begins building a binary tree data structure. In addition to the random makeup of the tree, a randomly selected scaling and sizing algorithm ensures that the visual representation of the tree never looks the same. During the drawing phase, when each node is created, a randomly chosen generated image of handwritten text will be placed next to it. Also, for each tree branch that is created, a randomly selected symbol is placed in the middle of the branch. This could be extended to use almost any shape, drawing, or another type of symbols or pictures. As noted previously, deformations may be applied not only to the handwritten images but to the tree itself to further discourage machine learning.

To complete the generation of our CAPTCHA, the program selects a node, or set of nodes in the rendered tree at random about which to ask the question. Once the tree and proper placement of images has been completed, a question will be selected at random from a list of potential questions about the tree such as “Which {word} is connected to {some other word} by a {shape}? (Figure 1). In all cases the user must quickly scan through and interpret all handwritten samples in the tree to correctly answer the question. Alternatively, the user may be asked to name two words that are connected. For machines to break our CAPTCHA, much more than one handwritten image would need to be interpreted. In addition, to solve our CAPTCHA machines would need to segment out the various objects in our CAPTCHA, including tree elements, shapes or other graphics, and handwritten images. Thus with our CAPTCHA we exploit the open problem in document image analysis of analyzing, segmenting and recognizing the various elements in a digital document or image [5]. To complete the generation of the test, the handwritten images and their truth words (correct answers) are passed to the verifier. Upon challenge submittal, the user response is verified and the application determines whether the user passes or fails. If the user passes by interpreting all of the letters in the handwritten image correctly, they would then be given access to the Web resource in question, otherwise they would be given another different challenge. Since the difficulty level of recognition needed to answer the question is greater and also has a higher level of randomness, our tree-based handwritten CAPTCHA poses more difficulty for machines, while still remaining simple for human users based on their cognitive abilities.

4. HIP System Evaluation

We have designed the tree-based handwritten HIP system as a challenge-response protocol for Web security. Experimental tests on word legibility have been conducted with human subjects and state-of-the-art handwriting recognizers. We have tested large sets of images on machines. To make it a fair test for machines, we have assisted the word recognizers with lexicons that contain all the truth words of the test images. For testing we used scanned handwritten image samples of US city names which we had readily available from postal applications, in order to provide samples corresponding to a known, finite lexicon (size 40,000, roughly the number of US city names) to help machine recognition, as well as synthetically generated samples. In reality, in actual applications such as our CAPTCHA having no context-specific dictionary, the

number of entries in the lexicon will be much larger which will affect recognition accuracy drastically, as indicated by researchers [13, 32].

4.1 Machine Testing

The handwritten CAPTCHAs have been tested by several state-of-the-art handwriting recognizers (Word Model Recognizer (WMR), Character Model Recognizer (CMR), and Accuscript (HMM)) [14, 31]. We have tested several sets of images using human-written scanned samples, synthetically generated samples, and tree-based handwritten images. Transformations were applied to human-written and scanned image samples based on the Gestalt principles and Geon theory. Several sets, each of them with over 4,000 handwritten city name images, were used, one set for each transformation. Parameter values for transformations were randomly chosen and successively applied to the handwritten word images. The individual transformations we were concerned with were less vs. more fragmentation, empty letters, displacement, mosaic effect, adding jaws, arcs, or extra strokes, occlusion by circles, occlusion by waves (white vs. black), vertical or horizontal overlap, and overlap of different words. We note that with the exception of occlusion by circle transformation, machine recognition rates were low for Gestalt-based transformations, even though the lexicon used to help machine accuracy was of a relatively small size. We observed that for occlusions by circle, machine recognition could be affected if the transformation did not adequately affect the foreground. We encountered overall accuracy of 5.74% for WMR, 1.21% for Accuscript and 3.8% for CMR, with accuracies approaching 0% for individual transformations such as letter fragmentations, overlaps, and adding extra strokes, when recognizers were aided by a relatively small lexicon of 4,000 words. On the other hand, these presented the least difficulty for human subjects, based on the Gestalt laws of closure and continuity [23, 24].

We also tested 300 synthetic handwriting samples that were generated corresponding to US city names, US states and world-wide countries. Similar Gestalt and Geon-motivated transformations were applied. For these images we saw an accuracy rate of only 1.00% for WMR, 0.7% for Accuscript, and 0.3% for CMR. This extremely low machine recognition rate even with a small word set and a provided small lexicon suggests that synthetic handwritten images are an excellent choice for generating infinitely many CAPTCHAs.

To the best of our knowledge, there is not yet any commercial program which can fully interact with or decipher our tree-based handwritten CAPTCHA. As we have noted previously, our use of tree structures is motivated by the fact that they can provide complexity for machine recognition beyond interpretation of handwritten images. In addition, we leverage open problems in graphics recognition as well as in the wider field of Document Analysis and Recognition. In our tree structure, symbols have no intrinsic meaning outside of our CAPTCHA. Thus, segmenting parts of our drawing would not assist in solving the combined CAPTCHA since the task of handwriting interpretation would still remain difficult. Our drawing only makes sense as a complete entity which requires the interpretation of symbols with no inherent meaning on top of interpreting handwritten transformed images with their previously mentioned difficulties for machines.

4.2 Usability Testing

Our usability testing focused on understanding the viability of our CAPTCHA both from a user experience perspective and based on how often users were able to interpret our CAPTCHA. A key area of focus was on determining whether our tree structure in combination with handwritten images presented any additional difficulty as compared to a handwriting only CAPTCHA. User tests were conducted both for handwritten images alone and for our tree-based handwritten CAPTCHA.

To test handwritten images alone that were transformed according to Gestalt and Geon principles, random sets of 90 images were given to be recognized by 9 volunteers. The test consisted of 10 handwritten word images for each of the 9 types of transformations. For the purposes of testing, to ensure human results could be fairly compared to machine results, images were chosen at random from transformed US city name images. The actual application will feature virtually infinite-many different synthetically generated and transformed word images with an unrestricted domain to foil machine recognition. We note that most of the human errors came from poor original images rather than being related to the transformations applied. Success rates for humans averaged 80%. As noted earlier, we believe that through a careful process of parameter selection and good quality starting samples, which can best be guaranteed by using synthetic samples, we can achieve a higher success rate for human recognition for our images. We have also compared human recognition on a set of human

handwritten US city name images to a set containing 79 synthetic US city name, state or country name images automatically generated by our handwriting generator program. Similar high human recognition of 80% or better for both human-written and synthetic image sets suggests that synthetic images pose no additional problems to humans.

Our first round of user studies on our tree-based handwritten CAPTCHA included 15 volunteer graduate and undergraduate students. Approximately 30% of the volunteers were non-native English speakers which suggests that our CAPTCHA may be useful to a large audience. Subsequent tests will be run with a larger set of participants and tests featuring a larger set of words. During our initial test, 190 challenges were completed and a series of survey questions were administered on a volunteer basis to all participants. Each participant was asked to take 20 tree-based handwritten CAPTCHA challenges. The tests were self-administered and were taken at a testing Web site. Participants were given only very basic information on the concept of CAPTCHA and no prior knowledge of the field was assumed. Users were asked to solve the CAPTCHA presented (Figure 1) and to rate each CAPTCHA on a scale of 1-5, with 1 being "least difficult" and 5 being "most difficult". At the end users were asked to provide general comments about our CAPTCHA as well as responses to specific questions about their Web usage and exposure to CAPTCHAs.

Users were able to interpret the tree-based handwritten CAPTCHA 80.6% of the time which was no less often than the 80% for handwritten CAPTCHA trials. The most common rating given for the tree-based handwritten CAPTCHA trials was 2, although the presence of trials rated as 4 or 5 made the average (2.8) slightly higher. We have observed that generally the samples given a higher rating were those with poor image quality from scanned samples. We feel that the average rating is acceptable given the current sample set and will only be improved with a cleaner, all synthetic set to be used in the next round of testing. One interesting observation is that in many cases where users rated an image as a 4 or 5, they were still able to correctly guess the image. We believe that recognizing additional elements in the tree allowed users to fill in the blanks and interpret images that they may not have otherwise been able to read. Several general comments collected in the survey support this assumption. The ability to guess words even when they were hard to read once again highlights the importance of human perception factors involved in reading, including those of local context and Gestalt and Geon principles. The same assistance provided to humans by the tree structure provided more difficulty for machines due to problems in graphics segmentation of complex multi-layer images. Of the individual transformations, more correct answers were for word images transformed using overlaps, mosaic effect and extra strokes.

5. Conclusions and Future Work

This paper furthers the work on handwritten CAPTCHA [22, 23, 24] and provides insights into the fields of AI, Handwriting and Graphics Recognition. A new approach is presented for a HIP system that combines handwritten word images in a randomly generated tree-based structure with randomly selected test questions. Our approach leverages currently superior human ability over machines in interpreting graphics and reading unconstrained handwriting, especially when various transformations such as fragmentations or occlusions are applied. The Gestalt laws of perception and Geon theory and the weaknesses of handwriting recognizers were closely studied to determine the most effective transformations to apply to keep images legible for humans while making them difficult to read by machines. We add the novel element of a randomly drawn tree structure with randomly drawn node elements in addition to handwritten images to further leverage human cognitive strengths over machines.

Experimental results show a high degree of human success even with inconsistent quality scanned handwriting samples, and user feedback has also been largely positive indicating that our CAPTCHA is human-friendly. At the same time, our tests using state of the art recognizers prove that machine success rates with the same tests are low to non-existent. Testing both real human handwriting and synthetically generated samples allowed us to compare the results at machine and human level and conclude that synthetic handwriting is at least as good as real handwriting for CAPTCHA purposes. All these aspects indicate that our CAPTCHA can successfully be used to protect online services as a viable alternative solution for Cyber security. Additionally, our CAPTCHA provides fertile ground for work on important problems in other areas such as AI, Image Analysis, Machine Learning, Security, etc., and invite researchers to work on breaking our CAPTCHA and further advance knowledge in those fields.

We are considering several improvements to our application based on user feedback. We will run additional user tests using a larger set of non-domain

specific words and synthetic samples generated and transformed on the fly by our handwriting generator. A wider range of participants will be considered and metrics on human time to solve the challenges by transformation type collected. More testing will be conducted to understand the processing load that our application might present during real-time use by many concurrent users. We also plan to have researchers create custom attacks to understand any potential vulnerabilities of our approach or will release our application to the public using some of our university online services in order to further understand how machines and a wider set of human users might interact with it. Combining handwritten text images with images of objects is another possible extension for our CAPTCHA. Last but not least, an alternative CAPTCHA for visually impaired users will be considered as well.

6. References

[1] Baird, H. S. and Popat, K. Human Interactive Proofs and Document Image Analysis. In Proc. IAPR Workshop on Document Analysis Systems, 2002.

[2] Biederman, I. Recognition-by-components: A theory of human image understanding. Psychological Review, 94, 2, 115-147, 1987.

[3] Biederman, I. and Blickle, T. The perception of objects with deleted contours. Unpublished manuscript, 1985.

[4] Biederman, I. and Gerhardstein, P.C.Recognizing Depth-Rotated Objects: Evidence and Conditions for Three-Dimensional Viewpoint Invariance. Journal of Experimental Psychology: Human Perception and Performance, 19, 1162-1182, 1993.

[5] Chaudhuri, B. Digital Document Processing: Major Directions and Recent Advances. Springer London, 2007.

[6] Chellapilla, K, Larson, K, Simard, P and Czerwinski, M. Designing Human Friendly Human Interaction Proofs (HIPs). In Proc. CHI 2005, ACM Press, 711-720, 2005.

[7] Chellapilla, K. and Simard, P. Using Machine Learning to Break Visual Human Interaction Proofs (HIPs). Advances in Neural Information Processing Systems, 17, 2004.

[8] Chew, M. and Baird, H.S. BaffleText: A Human Interactive Proof. In Proc. 10th IS&T/SPIE Document Recognition and Retrieval Conference, 2003.

[9] Freyd, J.J. Dynamic Mental Representation. Psychological Review, 94, American Psychological Assoc, 427-438, 1987.

[10] Gimpy web site: http://www.captcha.net/captchas/gimpy/

[11] Golle, P. Machine Learning Attacks Against the Asirra CAPTCHA. Proc. CCS '08, ACM, New York, 535-542, 2008.

[12] Goodman, J., Cormack, G.V., and Heckerman, D. Spam and the Ongoing Battle for the Inbox. Commun. ACM, 50:2, ACM, New York, 25-33, 2007.

[13] Govindaraju, V., Slavik, P. and Xue, H. Use of Lexicon Density in Evaluating Word Recognizers. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24, 6, 789-800, 2002.

[14] Kim, G. and Govindaraju, V. A Lexicon Driven Approach to Handwritten Word Recognition for Real-Time Applications. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19, 4, 366-379, 1997.

[15] Koffka, K. Principles of Gestalt Psychology. New York: Harcourt, Brace and Company, New York, 1935.

[16] Madhvanath, S. and Govindaraju, V. The Role of Holistic Paradigms in Handwritten Word Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23, 2, 149 – 164, 2001.

[17] Mori, G. and Malik, J. Recognizing objects in adversarial clutter: breaking a visual CAPTCHA. Proc. Computer Vision and Pattern Recognition, 1, I-134-I-141, 2003.

[18] Plamondon, R. and Srihari, S.N. Online and Off-Line Handwriting Recognition: A Comprehensive Survey., IEEE Transactions on Pattern Analysis and Machine Intelligence, 22, 1, 63-84, 2000.

[19] reCAPTCHA project – a CAPTCHA implementation. http://www.recaptcha.net/

[20] Rice, S., Nagy, G., and Nartker, T. Optical Character Recognition: An Illustrated Guide to the Frontier, Kluwer, Dordrecht, 1999.

[21] Rui, Y. and Liu, Z. ARTiFACIAL: Automated Reverse Turing Test Using Facial Features. Proc. MM '03, ACM, New York, 295-298, 2003.

[22] Rusu, A. and Govindaraju, V. Handwritten CAPTCHA: Using the difference in the abilities of humans and machines in reading handwritten words. In Proc. Ninth International Workshop on Frontiers in Handwriting Recognition, 226-231, 2004.

[23] Rusu, A. and Govindaraju, V. Visual CAPTCHA with Handwritten Image Analysis. Human Interactive Proofs: Second International Workshop, HIP 2005, Bethlehem, PA, USA, May 19-20, 2005: Proceedings, Springer, 42-51, 2005.

[24] Rusu, A. and Govindaraju, V. A Human Interactive Proof Algorithm Using Handwriting Recognition. In Proc. Eighth International Conference on Document Analysis and Recognition, 2, 29, 967-971, 2005.

[25] Rusu, A., Midic, U and Govindaraju, V. Synthetic Handwriting Generator for Cyber Security. In Proc. 13th Conference of the International Graphonomics Society, 2007.

[26] Senior, A.W. and Robinson, A.J. An Off-Line Cursive Handwriting Recognition System. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20, 3, 309-321, 1998.

[27] Srihari, S.N. and Kuebert, E.J. Integration of Hand-Written Address Interpretation Technology into the United States Postal Service Remote Computer Reader System. In Proc. Fourth International Conference on Document Analysis and Recognition, 2, 892-896, 1997.

[28] Turing, A. Computing machinery and intelligence. Mind 59, 236, 433-460, 1950.

[29] von Ahn, L., Blum, M., Hopper, N. and Langford, J. CAPTCHA: Using Hard AI Problems for Security. LNCS, Springer Berlin / Heidelberg, Vol. 2656, 294-311, 2003.

[30] von Ahn, L., Blum, M., and Langford, J. Telling humans and computers apart automatically. Communications ACM 47, 2, 56-60, 2004.

[31] Xue, H. and Govindaraju, V. A stochastic model combining discrete symbols and continuous attributes and its application to handwriting recognition. International Workshop of Document Analysis and Systems, 70-81, 2002.

[32] Xue, H. and Govindaraju, V. On the dependence of handwritten word recognizers on lexicons. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24, 12, 1553-1564, 2002.

[33] Yan, J. and El Ahmad, A.S. A Low-cost Attack on a Microsoft CAPTCHA. Proc. CCS '08, ACM, New York, 543-554, 2008.

[34] Yan, J. and El Ahmad, A.S. Usability of CAPTCHAs Or usability issues in CAPTCHA design. *Proc. SOUPS 2008,* ACM, New York, 44-52, 2008.

[35] Yan, L. and Wenyin, L. Engineering Drawings Recognition Using a Case-based Approach. Proc. Seventh Int'l Conf. on Document Analysis and Recognition, IEEE, Washington, D.C., 1-5, 2003.