EVALUATING A SEGMENTATION-RESISTANT CAPTCHA INSPIRED BY THE HUMAN VISUAL SYSTEM MODEL

I. M. KHAN, I. K. M. USHAMA AND O. O. KHALIFA

Electrical and Computer Engineering Department, Faculty of Engineering, International Islamic Univ. Malaysia (IIUM), Jalan Gombak, 53100 Kuala Lumpur, Malaysia

Email: imranmoezkhan@gmail.com

ABSTRACT: Visual CAPTCHAs are widely used on the Internet today as a means of distinguishing between humans and computers. They help protect servers from being flooded by requests from malicious scripts. However, they are not very secure. Numerous image processing algorithms are able to discern the characters used in the CAPTCHAs. It has been suggested that CAPTCHAs can be made more secure if they are distorted in ways that makes segmentation difficult. However, out of all the reviewed distortions present in current CAPTCHAs there are none that allow for a high level of segmentation difficulty. Furthermore, CAPTCHAs also need to be used by humans who may not find certain distortions tolerable. Thus, the problem of selecting a good distortion becomes a tradeoff between user acceptability and computer solvability. It is hypothesized in this paper that rather than using low-level image distortions, optical distortions based on the Gestalt laws of perception governing human visual system models should be applied. These distortions would ensure widespread user acceptability, and would be very difficult for computers to solve. This paper aims to explore the feasibility of employing Gestalt-inspired distortion in CAPTCHAs by first implementing a CAPTCHA cracker and then evaluating the performance of some manually generated Gestalt CAPTCHA's against some existing CAPTCHAs.

ABSTRAK: CAPTCHAs Visual digunakan secara meluas di Internet hari ini sebagai satu cara untuk membezakan antara manusia dan komputer. Mereka membantu melindungi server komputer daripada dibanjiri oleh permintaan daripada skrip yang berniat jahat. Walau bagaimanapun, server-server tersebut masih tidak selamat. Algoritma pemprosesan imej pelbagai mampu untuk membezakan watak-watak yang digunakan dalam CAPTCHAs. Ia telah dicadangkan bahawa CAPTCHAs boleh dibuat lebih selamat jika ia dicacatkan sehingga proses segmentasi adalah sukar. Pada waktu ini, semua perubahan yang dibuat di dalam CAPTCHAs masih belum boleh menghasilkan tahap kesukaran segmentasi yang tinggi. Tambahan pula, CAPTCHAs juga perlu digunakan oleh manusia yang tidak boleh menerima kecacatan dalam sistem CAPTCHAs. Oleh itu, masalah memilih perubahan CAPTCHAs yang bagus adalah di antara menerima kecacatan yang dapat diselesaikan oleh komputer dan juga diterima oleh pihak pengguna. Dalam kertas kerja ini adalah dicadangkan untuk tidak menggunakan herotan imej peringkat rendah, tetapi mengunakan ilusi optik yang berdasarkan undang-undang gestalt yang mengawal sistem visual manusia. Perubahan imej ini akan memastikan penerimaan pengguna yang meluas, dan meyukarkan usaha-usaha komputer untuk menyelesaikannya. Karya ini bertujuan untuk meneroka

kemungkinan mengunakan ilham perubahan gestalt untuk menghasilkan penyelesaian CAPTCHAs dan kemudian menilai prestasi CAPTCHA.

***KEYWORDS**: CAPTCHA, character recognition, image processing.*

1. INTRODUCTION

Human Interactive Proofs (HIP's) are often used by online servers to determine whether a request for transaction originates from a human or a computer, and form an integral part of online computer and network security nowadays. An HIP usually consists of a challenge or test that is designed such that only a human can pass it. The purpose of an HIP is to filter out requests that originate from malicious bot programs on the internet that would seek to cripple a server by bombarding it with thousands of requests a minute leading to a possible Denial of Service (DoS) attack. HIPs deployed at a major free email hosting server have been found to reduce exactly such traffic generated by spamming bots by 19 % [1, 2].

However, developing a standard HIP has proved to be quite a difficult task. Any challenge for an HIP can be selected so long as it is selected to be a 'hard' artificial intelligence (AI) problem. However, 'hard' AI problems are both numerous and may be inappropriate for deployment in a global environment such as the World Wide Web. It is important to remember that users come from a variety of backgrounds and may not recognize the objective of the challenge equally easily, or may find it too tedious to perform, thus abandoning it. This paper relies on the idea of using CAPTCHAs as HIPs, and would suggest a distortion for CAPTCHAs that is not currently employed but which is hypothesized to be far more efficient. This efficiency is tested in this paper by implementing a CAPTCHA cracker and then evaluating existing CAPTCHAs, as well as the suggested CAPTCHA via the cracker. This paper is arranged as follows: section 2 presents a background to CAPTCHAs and reviews some related work; section 3 gives an overview of some relevant Gestalt laws; section 4 provides implementation details of the CAPTCHA cracker; section 5 provides some results and discussion; and section 6 concludes.

2. BACKGROUND AND RELATED WORK

CAPTCHA, which stands for "Completely Automated Public Turing test to tell Computers and Humans Apart", is the most widely deployed Human Interaction Proofs (HIP) due to its simplicity and universality [1]. Introduced by Luis von Ahn in 2000, a CAPTCHA challenge presents a user with graphical text and requests the user to identify the characters used in the picture. However, image processing and computer vision algorithms have reached a point where it is not difficult for a computer to identify text in images any longer. As such there is a need to make CAPTCHAs more secure by increasing their difficulty to prevent them being 'cracked' by hackers. However, a tradeoff must be made between CAPTCHA difficulty and CAPTCHA usability as shown in Fig. 1.

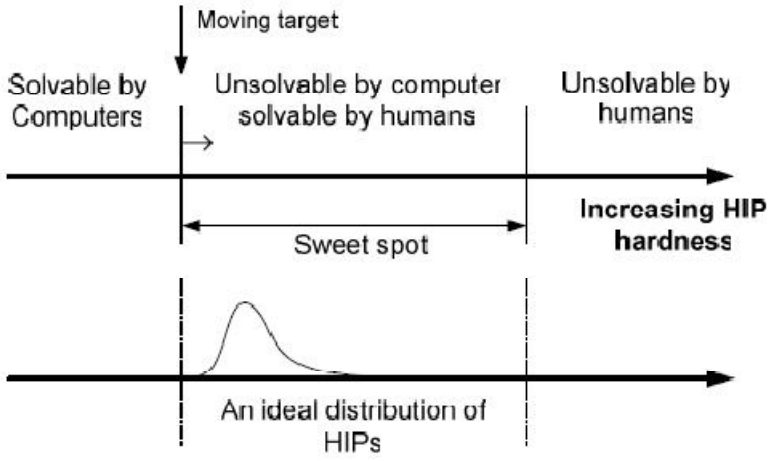

Fig. 1: Theoretical distribution of CAPTCHA difficulties [1].

It can be noted that, 'good' CAPTCHAs are those which fit the design requirements of being:

1) Difficult for computers to solve.
2) Easy for humans to solve.

Recent research is aiming to develop distortions that make segmentation of the image difficult, as segmentation is the most computationally expensive aspect of object/character recognition. As each CAPTCHA is unique and has different distortion parameters, it is difficult to design one robust system that manages to solve numerous different CAPTCHAs. Most segmentation algorithms are specifically formulated for particular CAPTCHA distortions. But, if a CAPTCHA image can be correctly segmented by an algorithm, it is usually considered 'solved', as the only next step that needs to be performed is that of classification.

The following sub-sections review the types of distortions currently employed in CAPTCHAs, their difficulty for computer solution, and their acceptability amongst users.

2.1 Background/Font Color

CAPTCHAs in which only background color is varied for distortion are extremely simple to implement and also solve [3]. Examples of CAPTCHAs that rely largely on background color variations to provide distortion are shown in Fig. 2.

Fig. 2: Examples of color-only distortions [3].

2.2 Linear Transformations

Linear transformation distortions in CAPTCHAs include translation, rotation and scaling. Chellapilla [1] conducted a user study to determine the acceptability of different distortions in CAPTCHAs. It has been found that the abovementioned linear transformation distortions are easily solved by humans with 99 % accuracy rate for maximally translated and rotated text, and 98 % accuracy rate for maximally scaled text.

However, although linear transformation distortion is shown to be widely accepted by users and also easy to implement, it is also the easiest to crack – with robust algorithms too. Ponec [3] investigated the use of affine moment invariants to eliminate the effects of translation, rotation and scaling in two different CAPTCHAs. Their use of 4^{th} order moment invariants led to 88 % accuracy of CAPTCHA solution over a span of 1000 test images.

2.3 Warping

Warping is a type of distortion produced when an image is subjected to a continuous displacement field. Warping in an image can be applied at two levels: globally and locally.

- Global Warp:

In global warp, the deformation of a character depends on the deformation of neighboring characters (refer to Fig. 3). In this figure, the character M is warped where the deformation of one character affects the others also.

Fig. 3: Example of global warp distortion [1].

- Local Warp:

In local warp, the displacement field within the characters themselves. Thus, character deformations are independent of each other (refer to Fig. 4). The deformation caused is from the original M character and not from the neighbor.

Fig. 4: Example of local warp distortion [1].

Chellapilla's [1] user study shows that human recognition rates for global and local warps are quite high up to a certain level of warping. However, Yan [4] and Moy [5] both develop algorithms that can easily segment a CAPTCHA with local and global warping. In particular, Yan [4] demonstrates the use of a simple, novel segmentation algorithm coupled with a feature matching classifier that is able to achieve a 99 % success rate on

their test set. Using template matching, a common confusion matrix and certain geometrical properties of the characters, Moy [5] is able to achieve 78 % accuracy on a blind testing database of 736 images. Yan [4] and Moy [5] thus are able to demonstrate their algorithms in acceptable processing of between 20 ms to 7 seconds.

2.4 Arc Noise

Arcs are a type of noise – rather than distortion – that can be added to a CAPTCHA. CAPTCHA's arc-noise can be classified into several types: thick intersecting arcs, thin intersecting arcs, thick non-intersecting arcs and thin non-intersecting arcs. Chellapilla [1] demonstrates that arcs which intersect with the characters are not tolerated well if they are thick.

However, Yan [6] develops a color-filling algorithm to successfully segment and break the Microsoft CAPTCHA at a success rate of 92 % in the test set and performs very efficiently at an average computation time of slightly over 80 ms. Assuming a conservative recognition rate of around 95 % after the segmentation operation, Yan [6] estimates that the Microsoft CAPTCHA can be solved more than 60 % of the time.

3. RELEVANT GESTALT LAWS

The human visual system (HVS) is made up of three major components:

1) The eye and surrounding tissues
2) The optic nerve
3) The visual cortex of the brain

Similar to computer vision, human visual processing is carried out in steps at increasing levels of abstraction and each of these components performs the different tasks that ultimately lead to the human visual sense. Models of the HVS attempt to explain the operation of these components at the different abstraction levels of visual processing. For example, models of the eye and retina explain image formation and focusing; models of the optic nerve explain the signal conditioning and pre-processing required before the image is interpreted in the brain; and models of the visual cortex explain visual perception.

Gestalt psychology refers to a set of principles that seem to govern higher human visual understanding and play an important role in visual perception. Developed in the early 20^{th} century, Gestalt psychological laws help to describe the little understood aspects of the human visual system such as optical illusions, automatic grouping and pattern recognition. These phenomenon arise due to the five Gestalt laws used in describing Gestalt systems [7]: Closure, Similarity, Proximity, Symmetry, and Smoothness. These are all illustrated in Fig. 5.

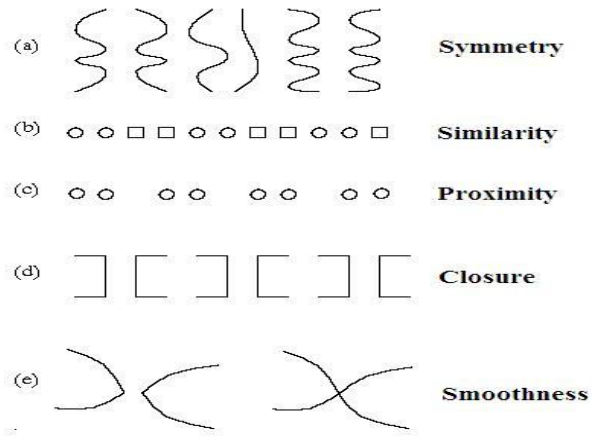

Fig. 5: Visual depiction of Gestalt Laws [7].

These laws are suggested in this paper as guidelines to producing CAPTCHA distortions that are natural for humans, but difficult for computers to solve.

4. IMPLEMENTATION OF CRACKER AND DISTORTION

A CAPTCHA cracker is implemented in MATLAB which uses a simple color and location based segmentation algorithm, followed by template creation from a training database of CAPTCHAs and finally uses correlation to perform template matching. The purpose of the training database is to make templates of the characters used in the CAPTCHA. The structure of the cracker is shown in Fig. 6.

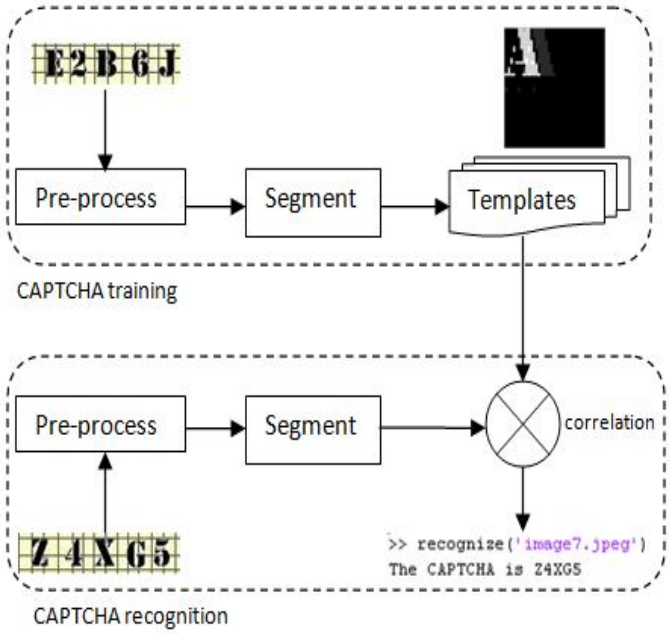

Fig. 6: Architecture of CAPTCHA cracker.

First, the preprocessing stage results in a binarized black and white image file. Next, the segmentation splits the CAPTCHA into its 5 individual characters. Thereafter the templates for individual characters are made from the training data using an averaging algorithm. These templates are then padded to create standardized sizes (refer to Fig. 6). The recognition depends on the correlation of the segmented testing CAPTCHA against the templates in the database.

Twenty-six training samples were taken to form the templates. It was made sure that each character available in all the CAPTCHA's had a template. Next, four different types of distortions were introduced into a total testing sample of 119 CATPCHAs. The four distortions relate to the Gestalt laws of Symmetry, Similarity, Proximity and Closure (refer to Fig. 7).

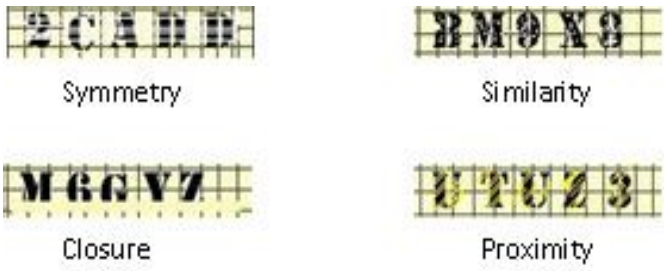

Fig. 7: Four distortions based on Gestalt Laws.

In order to test each type of distortion separately, the CAPTCHA images were first tested without the distortion, then tested with the distortion.

5. RESULTS AND DISCUSSION

The 119 CATPCHA samples were first all tested without distortion. This was done in order to record the accuracy of the CAPTCHA cracker before the samples were distorted. Next, the samples were divided into four different groups (Table 1), with a different Gestalt-inspired distortion being applied to each group.

Table 1: Testing sample for cracker.

Type of distortion	**Number of images**
Symmetry	32
Similarity	32
Closure	25
Proximity	30
TOTAL	119

The smallest number of samples was for the group chosen to be distorted by Closure distortion. The results of testing the cracker with these newly distorted samples, as well as

the undistorted samples are given in Table 2. Before distortion was applied, the CAPTCHA cracker implemented had an overall efficiency of 80.6 %. This dropped to an overall 50.4 % recognition rate after implementation of the distortions. In addition, it can be seen that the Closure type distortion has the greatest reduction in recognition rate.

For some of the testing samples the CAPTCHA cracker failed to recognize the CAPTCHA before and after the distortion. However, this failure of recognition may be for the same or different character after distortion. For example, a testing sample CAPTCHA spelled "ABCDE" may be wrongly identified as "BBCDE" before distortion, and "BBCDE" after distortion. In this case introducing the distortion did not make a difference. But, if the CAPTCHA was wrongly identified as "BBCDE" before distortion, and "BBBDE" after distortion, then this means that introducing the distortion made a difference in the error made by the CAPTCHA cracker.

Table 2: Identification results before and after distortions.

Type of distortion	**Correctly identified before distortion**	**Percentage of total**	**Correctly identified after distortion**	**Percentage of total**
Symmetry	28	87.5 %	18	56.25 %
Similarity	26	81.25 %	20	62.5 %
Closure	22	88 %	6	24 %
Proximity	20	66.667 %	16	53.33 %
OVERALL	96	80.6 %	60	50.4 %

This is a better means of evaluating the efficiency of the type of distortion rather than simply counting the number of failures before and after distortion. Table 3 provides a summary of these results. For example, in the case of Symmetry distortion, it is known from Table 2 that before distortion, 28 samples were correctly identified, and after distortion 18 samples were correctly identified. Thus, the distortion caused a difference in 10 correctly identified samples. However, out of the samples that failed to be identified both before and after distortion, there was one CAPTCHA sample that was identified differently (i.e. the cracker made a different mistake). Thus, the Symmetry distortion made a difference in a total of 11 CAPTCHA samples.

Table 3: Difference caused by distortions.

Distortion	**Total Differently Identified CAPTCHA's**	**Percentage**
Symmetry	11	34.37 %
Similarity	7	21.87 %
Closure	20	80 %
Proximity	6	20 %

From Table 3, it can be clearly seen that the Closure distortion was highly effective in introducing errors in the CAPTCHA cracker, and managed to confuse the CAPTCHA

cracker 80 % of the times causing it to produce a different result for both the correctly and incorrectly recognized CAPTCHAs. The least effective methods were Proximity and Similarity both of which offered only a marginal amount of increase in the error for the CAPTCHA cracker and should not be used for introducing Gestalt inspired distortions.

6. CONCLUSION

CAPTCHAs are amongst the most widely used HIPs on the Internet. Their working principle lie in distorting text characters in such a way that recognition becomes difficult for computers (< 0.01 % accuracy), but remains easy for humans (> 90 % accuracy) [1]. However, there is much interest now in developing image processing algorithms that can effectively identify the characters in CAPTCHAs despite the noise and distortion present [5]. These algorithms will not only have an impact on strengthening CAPTCHA's and consequently HIP's, but will also develop state-of-the-art optical character recognition (OCR) techniques [8]. Machine learning and image processing techniques have been used to 'break' CAPTCHA's quite effectively at very high character recognition rates (> 60 %) [5, 9, 10, 11].

This paper reviews currently deployed CAPTCHA distortions and the Gestalt properties of the HVS in the context of creating simple CAPTCHA distortions that are difficult for computers to solve. A CAPTCHA cracker was implemented in MATLAB that followed a correlation based template-matching algorithm to recognize the characters in a CAPTCHA. Twenty-six CAPTCHAs were randomly chosen to train the cracker. Next, 119 CAPTCHA samples were divided into four different groups to implement four different types of Gestalt-inspired distortions in order to determine whether the recognition problem became any harder for the computer. It was found that Gestalt-inspired distortions made recognition more difficult in all cases, especially when the distortion involved the principle of Closure.

Future work should involve investigating the distortion on more advanced recognition algorithms such as artificial neural networks, or support vector machines. Additionally, the prospect of automating the application of a Gestalt inspired distortion should be explored, as there are currently no techniques to implement them automatically.

REFERENCES

[1] Chellapilla K, Larson K, Simard P, Czerwinski M, (2005a) *"Designing human friendly interaction proofs,"* Proceedings of the 2005 Conference on Human Factors in Computing Systems, CHI 2005, Portland, Oregon, USA, April 2-7, 2005

[2] Hoque M, Russomanno D, Yeasin M, (2006) *"2D CAPTCHAs from 3D Models"*, IEEE South East Conference, Memphis, TN, USA.

[3] Ponec M, *"Visual Reverse Turing Tests: A False Sense of Security,"* 7th IEEE Information Assurance Workshop, Westpoint, New York, USA, 2006.

[4] Yan J, Ahmad SA, *"Breaking Visual CAPTCHAs with Naïve Pattern Recognition Algorithms,"* 23rd Annual Computer Security Applications Conference, 2007.

[5] Moy G, Jones N, Harkless C, Potter R,*"Distortion estimation techniques in solving visual CAPTCHA's,"* In Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, Vol. 2, 2004, pp. II-23-II-28.

[6] Yan J, Ahmad SA,*"A Low-Cost Attack on a Microsoft CAPTCHA,"* Proc. of 1^{5th} ACM Conference on Computer and Communication Security, Alexandria, Virginia, USA, 2008.

[7] Robert S., *"Cognitive Psychology"* 3rd Edition, Thomson Wadsworth Publishers, 2003.

[8] Simard P, Szeliski R, Benaloh J, Couvreur J, Iulian Calinov,*"Using Character Recognition and Segmentation to Tell Computer from Humans,"* Proceedings of the Seventh International Conference on Document Analysis and Recognition, August 03-06, 2003, 418 p.

[9] Chellapilla K, Simard P, *"Using Machine Learning to Break Visual Human Interaction Proofs (HIPs),"* Advances in Neural Information Processing Systems 17, Neural Information Processing Systems (NIPS'2004), MIT Press, 2004.

[10] Chellapilla K, Larson K, Simard P, Czerwinski M,*"Computers beat Humans at Single Character Recognition in Reading based Human Interaction Proofs (HIPs),"*. CEAS 2005 - Second Conference on Email and Anti-Spam, July 21-22, 2005, Stanford University, California, USA, 2005.

[11] Mori G, Malik J, *"Recognizing Objects in Adversarial Clutter: Breaking a Visual CAPTCHA,"* Proc. of Comp. Vision and Pattern Rec. (CVPR) Conf., IEEE Computer Society, vol.1, pages:I-134 - I-141, June 18-20, 2003.