Framework for Evaluation of Text Captchas

Achint Thomas
Yahoo! Labs
aothomas@yahoo-inc.com

Kunal Punera
RelateIQ
kunal.punera@utexas.edu

Lyndon Kennedy
Yahoo! Labs
lyndonk@yahoo-inc.com

Belle Tseng
Apple Inc.
belletseng@gmail.com

Yi Chang
Yahoo! Labs
yichang@yahoo-inc.com

ABSTRACT
Interactive websites use text-based Captchas to prevent unauthorized automated interactions. These Captchas must be easy for humans to decipher while being difficult to crack by automated means. In this work we present a framework for the systematic study of Captchas along these two competing objectives. We begin by abstracting a set of distortions that characterize current and past commercial text-based Captchas. By means of user studies, we quantify the way human Captcha solving performance varies with changes in these distortion parameters. To quantify the effect of these distortions on the accuracy of automated solvers (bots), we propose a learning-based algorithm that performs automated Captcha segmentation driven by character recognition. Results show that our proposed algorithm is generic enough to solve text-based Captchas with widely varying distortions without requiring the use of hand-coded image processing or heuristic rules.

Categories and Subject Descriptors
K.6.5 [**Management of Computing and Information Systems**]: Security and Protection – *authentication, unauthorized access.*

General Terms
Measurement, Security, Human Factors

Keywords
Captcha, human interactive proofs, bots

1. INTRODUCTION
Text-based Captchas are popular since recognition of degraded, noisy, distorted text with background clutter is a task that humans perform with relative ease compared to bots. Given the widespread use of text-based Captchas, it is surprising that there are few works in literature that describe strategies for the design of Captchas that maximize the gap between human and bot solving rates. Most Captchas are designed through intuitive rules of thumb and validated via heuristic experiments. This has led to the development of many successfully attacks by special-purpose bots [1,2,3,4]. The only work to systematically compare human and bot solving rates is [2] where the authors only tested recognition performance on pre-segmented, single characters with single distortions applied. The published attacks on CAPTCHAs [1,2,3,4] have taught us that segmentation is harder than recognition; indeed, [2] reports that bots are better than humans at the task of recognizing distorted single characters. However, we know from the continued popularity of Captchas on the Web that this is not true of recognition accuracies of humans and bots on complete Captchas. Our goal in this work is to benchmark the human and bot recognition performance and rigorously study what differentiates the solving abilities of humans and bots on complete Captchas where the subjects have to solve the segmentation task.

2. METHODOLOGY
2.1 Testing on Captcha Images
We conducted a survey of the existing major past and existing Captchas and decomposed the types of distortions found in them into six major classes. The classes identified agree well with existing literature [1,2].

Table 1: Various image distortions classes and their presence in existing and past CAPTCHAs.

	Affine Transforms	Kerning Overlap	Local Wrap	Global Wrap	Spurious Foreground	Missing Ink
Yahoo/Google Wikipedia	x	x	x	x		
Reddit	x			x	x	
MSN/eBay Baidu/CNN	x	x	x	x	x	
reCaptcha MegaUpload	x	x		x		x
mail.ru	x	x			x	
captcha.net	x					x
digg.com slashdot	x				x	

We then constructed an end-to-end Captcha generation system where all these distortions could be included with varying levels of hardness. These distortion classes and their presence in Captchas of major web services are summarized in Table 1. Using this distortion framework allows us to construct Captchas of arbitrary hardness that conform to actual Captchas used in practice in the industry.

2.2 Generic Captcha Solver
We develop a generic Captcha solver that uses high-precision character recognition to drive the segmentation process. We assume that an attacker has available (i) a large number of sample Captcha image instances, (ii) the text solutions to these instances, and (iii) per-character segmentation boundaries (the left and right boundary locations) for each character in an instance. A motivated attacker can easily achieve these requirements. At its core, our Captcha segmentation strategy is similar to classical image

template matching in which the goal is to detect the presence of some object by searching over the entire image using an exemplar of the object to be found. We rely on high-precision character recognition to drive the segmentation process. Instead of trying to explicitly identify high-confidence character segmentation boundaries for the individual characters in a Captcha, an automated solver can try various candidate character segmentation boundaries, compute the confidence of there being a complete character in that segment, and pick the sequence of boundaries that yields the highest confidence solution. The segmentation is implemented as a dynamic programming search through the various candidate character segment sequences possible for an image subject to pruning criteria that limits the smallest and largest segment widths possible. Segment widths less than the expected width of the narrowest character or larger than the expected width of the widest character in the dataset are ignored when searching for the solution.

3. EXPERIMENTS

For all the experiments that follow, we varied that feature across a range of parameters and tested the recognition accuracy of humans and bots on samples at each distortion level. Only one feature was varied at a time, and the remaining features were held at a nominal value. We selected a nominal parameter value for each distortion feature which corresponds to a very low, but non-zero distortion effect.

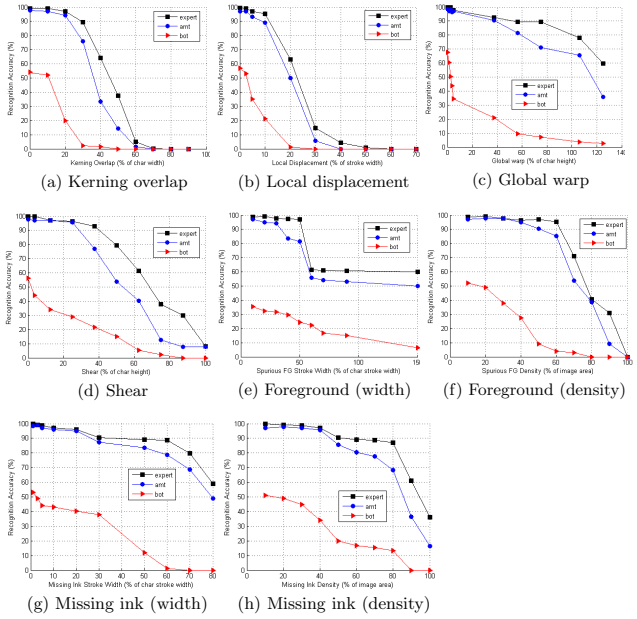

Figure 1: Recognition accuracy of humans and bot as difficulty parameter is varied for various distortion classes

A total of 10,000 Captcha instances were presented separately to *Expert* subjects, *AMT* subjects, and the bot solver. There were 20 distinct *Expert* subjects and 203 distinct *AMT* subjects who independently solved the 10,000 instances. The plots in Figure 1 show the performance of humans and the bot on the various distortion classes. The results for all distortions show that at the lower distortion levels, both *EXPERT* and *AMT* subjects perform comparably well while recognizing Captchas. Similarly, when the distortion levels are sufficiently high, both *EXPERT* and *AMT* subjects perform equally badly (see Figures 1(a),1(b), 1(d), and 1(f)). For mid-level distortion levels, there is a marked decrease (10-15% lower) in the recognition accuracy of *AMT* subjects compared to *EXPERT* subjects.

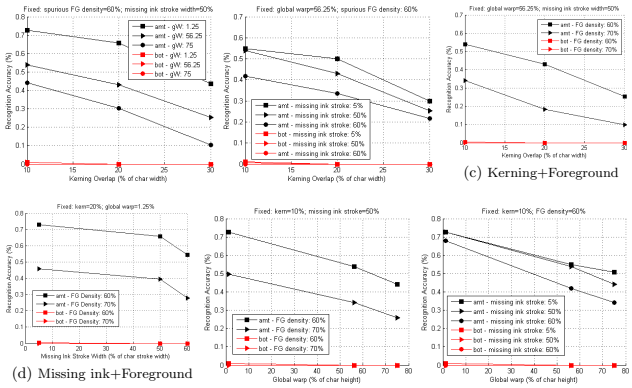

Figure 2: Recognition accuracy of humans and bot for various combinations of distortion classes

To further drive down bot recognition accuracy we can simultaneously vary multiple features. However, this will have a detrimental effect on human recognition as well. In order to test exactly how much this performance drop is, we setup the following experiments. First, we looked at the results for individual distortions and identified those classes for which the bot recognition accuracy is less than 10% when human recognition accuracy is greater than 70%. This yielded four classes: (i) Character Kerning, (ii) Global Warp, (iii) Spurious FG Density, and (iv) Missing Ink Stroke Width. We generated a set of 5,200 CAPTCHA instances by varying each unique pair of combinations of the four features identified above. When the parameters for any two features were being varied, the remaining two features were held fixed at constant values, which corresponded to acceptable levels of human recognition accuracy. The plots in Figure 2 show the recognition accuracies of *AMT* subjects and the bot solver as pairs of distortion features were varied.

4. CONCLUSION

In this work, we conducted an evaluation of human and bot performance on text-based Captchas. We identified a set of common Captcha image distortions by studying various existing and past Captchas, combined these distortions to construct a generic Captcha and conducted tests to understand what effect varying the strength of these distortions had on the recognition abilities of humans and bots. We presented (to the best of our knowledge) the first learning-based recognition-driven segmentation framework that can simulate other Captcha-specific solving attacks and that can be used for the purpose of testing the strengths of any newly developed Captcha technique.

5. REFERENCES

[1] Bursztein, E., Martin, M., and Mitchell, J., Text-based Captcha strengths and weaknesses, 18th ACM conference on Computer and communications security, pp. 125–138. ACM, 2011.

[2] Chellapilla, K., Larson, K., Simard, P., and Czerwinski, M., Computers beat humans at single character recognition in reading based human interaction proofs (HIPs), 2nd Conference on Email and Anti-Spam, pp. 21–22. Citeseer, 2005.

[3] Yan, J., and El Ahmad, A., Breaking Visual CAPTCHAs with Naive Pattern Recognition Algorithms, 23rd Annual Computer Security Applications Conference, 2007. ACSAC 2007, pp. 279–291. 2007.

[4] Yan, J., and El Ahmad, A., A low-cost attack on a Microsoft captcha, 15th ACM conference on Computer and communications security, pp. 543–554. 2008.