A Human Interactive Proof Algorithm Using Handwriting Recognition *

Amalia Rusu and Venu Govindaraju
Center of Excellence for Document Analysis and Recognition
University at Buffalo
Buffalo, NY, USA
{air2,govind}@cedar.buffalo.edu

Abstract

The recognition of unconstrained handwriting continues to be a difficult task for computers despite active research for several decades. This is because handwritten text offers great challenges such as: character and word segmentation, character recognition, variation between handwriting styles, different character size and orientation, no font constraints, the type of printing surface, as well as the background clarity. In this paper we explore the gap in the ability in reading handwritten text between humans and computers to propose solutions for security problems in Web Services. We present a new HIP algorithm that uses handwriting recognition task to distinguish between humans and computers. We propose methods to deform handwritten text images to make them indecipherable by computers and explore the cognitive factors that assist humans in reading and understanding. Experimental results on both humans and computers are presented and compared.

1. Introduction

Examples of handwriting recognition applications are processing mail addresses, interpreting bank checks, medical, census, tax forms, and more recently writer identification and signature verification [5, 8, 12]. In addition, our paper introduces a new practical application of handwriting recognition in security of Web services. We propose using the differences between humans and machine in abilities of reading handwritten text.

Although not the only option, one natural way to differentiate humans from computers is to use challenges that involve recognition of handwritten text. Making a computer perform the handwriting recognition task involves techniques from different areas, including pattern recognition, image processing, computer vision, artificial intelligence, language understanding, and psychology. Despite the research effort in all of these areas, handwriting recognition continues to be a challenge for computers even though it is a simple task for humans.

In this paper we design a HIP system that uses handwritten text images as (CAPTCHA) challenges (Figure 1). We identify some of the problems that current state-of-the-art OCRs face and describe how handwritten text images could serve as challenges. In order to expand the gap between human and machine, we automatically create handwritten text image challenges by altering visual features and use cognition laws to explain why humans are still able to correctly decipher them, but machines cannot.

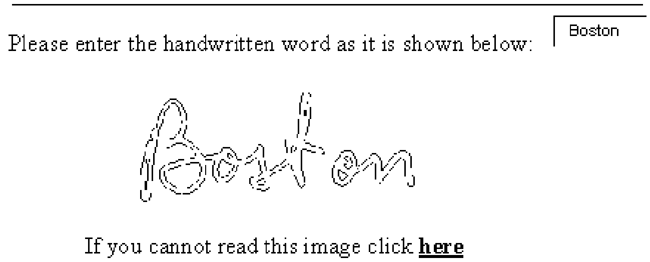

Figure 1. Example of HIPs interface and handwritten CAPTCHA challenge to confirm human registration.

2. Motivation and Previous Work

Human Interactive Proofs (HIPs) is a relatively new active research area that focuses on defending online services against abusive attacks. HIPs use a set of security protocols based on automatic tests (CAPTCHA) that virtually all humans can pass but current computer programs fail [1]. Previous research effort on HIPs and CAPTCHAs have taken advantage of the superior human ability in read-

* Research supported in part by Calspan-UB Research Center, Buffalo, NY.

ing machine printed text, as well as using speech and facial features recognition [2, 3, 7, 10, 13]. Several reading-based HIP systems are currently in use by AltaVista, Hotmail, Yahoo, Google, Ticketmaster, etc. A strong motivation for our research work is that some of the current machine printed text based HIPs have been systematically broken [*www.captcha.net*]. To the best of our knowledge, there is no handwritten text based HIP system used or developed so far. We propose a handwritten text based HIP system that exploits the gap between human and machine recognition of handwritten words and present it as a solution for securing online services against malicious automatic attacks.

3. Technical Approach

Our reading-based HIPs algorithm consists of the following steps for user authentication. First the user expresses an interest in being authenticated by the server, then the server generates a challenge in the form of a handwritten word image and issues it to the user. The user has to key in the right answer and return it to the server. Finally, the server verifies the user response and checks if it matches the right answer to either grant access to the user or reject the transaction.

The common sources of errors for OCRs are segmentation errors, such as over-segmentation or inability to segment, recognition errors, like confusing with a similar entry in lexicon, and image quality. Humans in general are better at segmenting handwritten text from an image with background noise, occlusions or fragmentations [9].

Human perception uses organizing principles called *Gestalt Laws*. Some of them are: the laws of closure, similarity, proximity, symmetry, continuity, familiarity and figure-ground distinction (Figure 2). We tend to complete a figure in the way it should be. If something is missing in an otherwise complete figure we will tend to close it, and eventually fill the gap (closure). We tend to group similar items together to see if they could form a larger object (similarity). Humans imagine things that are close together as belonging together (proximity).

Our approach is to remove features or add non-textual strokes and noise to a handwritten image in a systematic fashion based on Gestalt segmentation and grouping principles, in order to break machine recognition but preserve overall letter legibility and word recognition for humans.

Several transformations and their correspondence to perception laws are described here.

- Given the law of closure, proximity, and continuity, humans are able to correctly recognize handwritten text that contains fragmentation, empty letters, broken letters, or edgy contours (Figure 3).
- Given the law of closure, proximity, continuity, and symmetry, humans easily recognize an image that consists of a split in the horizontal direction and slightly displaced in opposite directions (Figure 4).

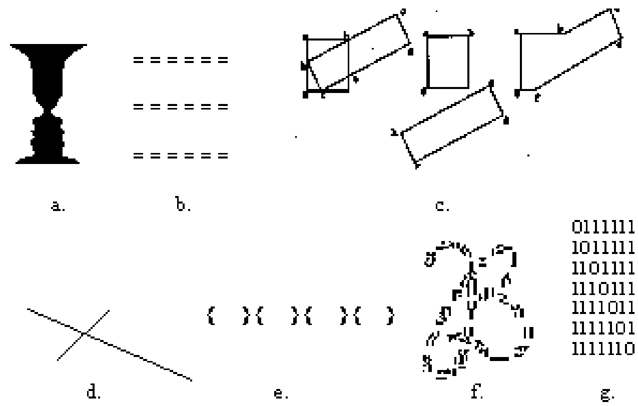

Figure 2. Image analysis using Gestalt Laws. Common examples for figure-ground (a), proximity (b), familiarity (c), continuity (d), symmetry (e), closure (f), and similarity (g).

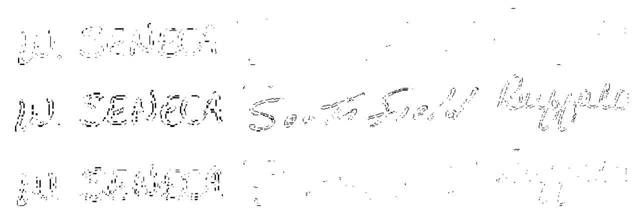

Figure 3. Examples of handwritten images where OCR systems fail. The truth words are: *W.Seneca*, *Southfield*, and *Buffalo*.

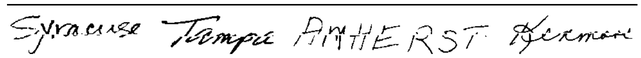

Figure 4. Examples of handwritten images where OCR systems fail. The truth words are: *Syracuse*, *Tampa*, *Amherst*, and *Kenmore*.

- Similar idea if there is a split image in multiple parts, either by a vertical, or horizontal line or by diagonals, and the parts are spread apart leaving empty spaces in between the original stroke, i.e. mosaic effect (Figure 5).
- Given the figure and ground distinction assumptions and the familiarity of letters, by adding occlusion using the same pixels as the foreground pixels, jaws, arcs, or lines, we do not add significant difficulties to the human reading (Figure 6).

Proceedings of the 2005 Eight International Conference on Document Analysis and Recognition (ICDAR'05)
1520-5263/05 $20.00 © 2005 **IEEE**

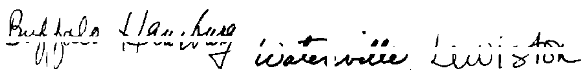

Figure 5. Examples of handwritten images where OCR systems fail. The truth words are: *Buffalo*, *Hamburg*, *Waterville,* and *Lewiston*.

Figure 6. Examples of handwritten images where OCR systems fail. The truth words are: *Binghamton*, *Lockport*, *Rochester*, and *Bradenton*.

The advantage of using handwriting for HIP systems is that most handwritten text challenges are uniquely problematic and usually more severe than problems encountered in recognizing machine-printed text. Segmentation is a complex and computationally expensive step which usually creates problems to recognition if not effectively solved.

By controlling the process of adding, deleting, or modifying strokes and character features we preserve the human recognition capabilities since gestalt principles assist humans in recognition process.

4. Experiments and Results

To test the handwritten CAPTCHA challenges, we used state-of-the-art word recognizers: Word Model Recognizer (WMR) and Accuscript [4, 6, 15]. WMR is a segmentation-based recognizer that treats each word as a model. The objective is to find the best match between a word in the lexicon and the image. The general processing stages are distinguished as: training, preprocessing, segmentation, feature extraction, and recognition. On the other hand, Accuscript is a grapheme-based recognizer, which extracts features from sub-characters such as loops, turns, junctions, arcs, without explicit segmentation. Both recognizers use static lexicons in the recognition process.

In our tests, we applied several image transformations on a set of about 4,100 handwritten city name images. We created several new sets of transformed images, one set for each deformation method previously described. We randomly choose some parameter values for our transformations and successively apply them on handwritten word images. We assume that a valid lexicon is provided and that it comprises of all the city names that we use in our tests. We ran tests on lexicons of size 4,000 and 40,000 - corresponding to the entire list of US city names.

Both humans and machine recognizers' accuracy was computed as the percentages of recognized images. The CAPTCHA tests are fully graded pass or fail, where pass is granted when all the characters of the word are correctly recognized, and fail otherwise.

The accuracy achieved by machine recognizers is presented in Table 1 for WMR and Table 2 for Accuscript recognizer. The most efficient transformations, based on our results, are letters fragmentations. The gestalt laws of closure and continuity hold strongly in this case, and humans easily fill the gaps or continue the characters that are broken apart.

One might expect low accuracy for handwriting recognizers when jagged strokes are added to the original images. Jagged strokes and arcs, as well as regularly spaced and sized graphics, or short drawings, can be misclassified as text and lead to segmentation failure.

For the splitting transforms, without losing the randomness of the method, we used middle cuts, however different locations for cuts in both horizontal and vertical direction can be considered, or different type of image padding can be randomly applied. Although we did preliminary tests varying some parameters, we did not record a comparison among them at this time, as it is not part of our current study.

Transformation	*WMR*	
	L = 4,000	*L = 40,000*
All Transformations	6.76%	3.09%
Empty letters	0.89%	0.38%
Small Fragmentation	0.00%	0.00%
High Fragmentation	0.48%	0.19%
Displacement	19.75%	10.27%
Mosaic	14.34%	6.42%
Jaws/Arcs	5.12%	1.33%

Table 1. The accuracy of WMR for current image transformations. A set of 4,127 images was tested for each kind of transformation using lexicons with size 4,000 and 40,000.

If we compare against other types of image deformations, some of the current methods perform better (e.g. letters fragmentations, or adding jaws). We compared with several other transformations such as adding lines, grids, arcs, background noise, applying convolution masks and special filters, using variable stroke width, slope, rotations, stretching, compressing (Table 3). Up to three such image transformations were previously applied on each test image [11]. Therefore, the current image transformations may be considered better than previous ones since we are using only one transformation per image and combining more than one will decrease the recognizers' accuracy further.

We administered tests for 9 voluntary students. The test consists of about 10 handwritten word images per each of

Transformation	Accuscript	
	L = 4,000	L = 40,000
All Transformations	3.60%	1.21%
Empty letters	0.06%	0.02%
Small Fragmentation	0.18%	0.16%
High Fragmentation	0.00%	0.00%
Displacement	8.84%	3.36%
Mosaic	8.99%	2.98%
Jaws/Arcs	3.58%	0.77%

Table 2. The accuracy of Accuscript recognizer for current image transformations. A set of 4,127 images was tested for each kind of transformation using lexicons with size 4,000 and 40,000.

Word Recognizer	Accuracy
WMR	9.28%
Accuscript	4.41%

Table 3. The accuracy of handwriting recognizers for other image transformations.

the 6 types of transformations previously described. The images were chosen at random from the sets of about 4,100 deformed images per transformation. The human subjects were relatively familiar with the words in the images since they are city names in the US. The tests suggest that human performance depends on context, and prior knowledge of the word provides the greatest advantage to human readers. Therefore, memory and word familiarity (Gestalt principles) have proven to be valid clues for humans. Usually, if the original handwritten sample was clean to begin with, after deformation it did not create problems for humans, but does for machines. However, if the original sample contains noise or was poorly written, then even the original image causes problem to both human and computer, even before deformation. Therefore, most of the human errors come from non-sensical original images rather than difficulties with the deformations applied on those images. The human results are presented in Table 4. The gap in the ability in recognizing handwritten text between humans and computers is illustrated in Figure 7.

The recognition of unconstrained handwriting is difficult because of the great variability in writing styles, spacing between words and lines, character sizes, and shape similarity. Different types of printing and background clarity also add to this challenge. On the other hand, the accuracy of handwriting recognizers is dependent upon the size and density of the lexicon, and mainly its availability [14, 16]. For our purpose and to give machines a fair shot, we con-

Transformation	# of Images	Accuracy
All Transformations	535	75.89%
Empty Letters	89	82.02%
Small Fragmentation	88	73.86%
High Fragmentation	90	74.44%
Displacement	89	78.65%
Mosaic	90	74.44%
Jaws/Arcs	89	71.91%

Table 4. The accuracy of human readers.

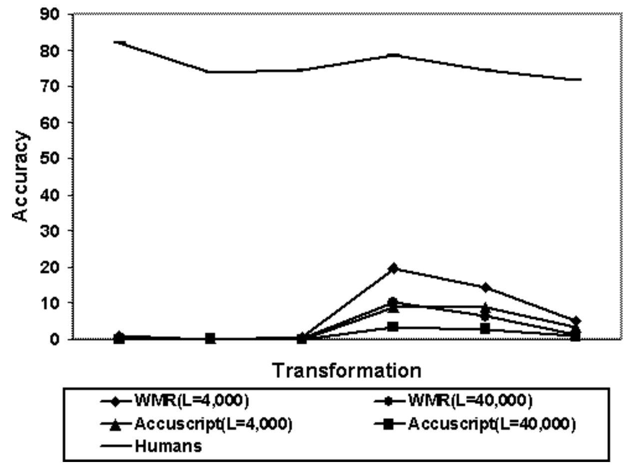

Figure 7. The gap in the ability in recognizing handwritten text between humans and computers per type of transformation. The transformations considered are empty letters, small fragmentation, high fragmentation, displacement, mosaic and jaws.

sider here lexicons containing all the truth words, and we did not consider increasing the lexicon size or density. However, for a general unconstrained example we should consider a large subset of the English dictionary, or use lexicon free approaches.

5. Summary and Future Work

Comparing humans and computers performance, we see that slightly deformed images do not pose problems to humans whereas the handwritten CAPTCHA images remain challenging for state-of-the-art recognizers. Our experiments revealed a considerable gap in abilities of reading handwritten word images between humans and computers, and unless more cognitive factors are included into the machine recognition steps, computers would fail when text is poorly written, fragmented or occluded.

In this paper we described new ways of creating Handwritten HIPs challenges. We analyzed the image transformations that use gestalt components and tested various sets of images on two advanced handwriting recognizers. The results obtained have been encouraging. We will reconsider some of the transformations and vary the parameters or use a combination of transformations in order to lower the recognizer accuracy. Generating manifold samples from handwritten words will be considered to automatically generate infinite-many (CAPTCHA) tests.

References

[1] H. Baird and K. Popat. Human interactive proofs and document image analysis. *Proc. IAPR 2002 Workshop on Document Analysis Systems*, August 2002.

[2] M. Chew and H. Baird. Baffletext: A human interactive proof. *Proc. SPIE-IST Electronic Imaging, Document Recognition and Retrieval*, pages 305–316, January 2003.

[3] A. Coates, H. Baird, and R. Fateman. Pessimal print: a reverse turing test. *Proc. IAPR 6th International Conference on Document Analysis and Recognition*, pages 1154–1158, September 2001.

[4] J. T. Favata. Character model word recognition. *Proc. Fifth International Workshop on Frontiers in Handwriting Recognition*, pages 437–440, September 1996.

[5] S. Impedovo, P. Wang, and H. Bunke. Automatic bankcheck processing. *Machine Perception and Artificial Intelligence*, World Scientific 28, 1997.

[6] G. Kim and V. Govindaraju. 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.

[7] G. Kochanski, D. Lopresti, and C. Shih. A reverse turing test using speech. *Proc. International Conference on Spoken Language Processing*, September 2002.

[8] S. Madhvanath, V. Govindaraju, V. Ramanaprasad, D. Lee, and S. Srihari. Reading handwritten us census forms. *Proc. Third International Conference on Document Analysis and Recognition*, pages 82–85, 1995.

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

[10] Y. Rui and Z. Liu. Artifacial: Automated reverse turing test using facial features. *Proc. The 11th ACM international conference on Multimedia*, November 2003.

[11] A. Rusu and V. Govindaraju. Handwritten captcha: Using the difference in the abilities of humans and machines to read handwritten words. *Ninth IAPR International Workshop on Frontiers of Handwriting Recognition*, pages 226–231, October 2004.

[12] S. Srihari and E. Keubert. Integration of hand-written address interpretation technology into the united states postal service remote computer reader system. *Proceedings of Fourth International Conference on Document Analysis and Recognition*, pages 892–896, August 1997.

[13] L. von Ahn, M. Blum, and J. Langford. Telling humans and computers apart (automatically) or how lazy cryptographers do ai. *Technical Report TR CMU-CS-02-117, Carnegie Mellon University*, February 2002.

[14] H. Xue and V. Govindaraju. On the dependence of handwritten word recognizers on lexicons. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(12):1553–1564, December 2002.

[15] H. Xue and V. Govindaraju. A stochastic model combining discrete symbols and continuous attributes and its applications to handwriting recognition. *International Workshop on Document Analysis and Systems*, pages 70–81, 2002.

[16] H. Xue, V. Govindaraju, and P. Slavik. Use of lexicon density in evaluating word recognizers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(6):789–800, June 2002.