The Ability Gap Between Human & Machine Reading Systems

Henry S. Baird
Xerox Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, CA 94304 USA
E-mail: baird@parc.xerox.com

Abstract

1 Who am I?

I am a Principal Scientist and manager of the Document Image Decoding research area at the Xerox Palo Alto Research Center.

My research interests include document image analysis, digital libraries, the design and analysis of algorithms, computational geometry, computer vision, combinatorial optimization, and software engineering.

In recent years I have focused on algorithms and systems architecture for extremely versatile printed-page readers that are easily retargetable to new applications, including non-English languages and non-Latin writing systems. A current major interest of mine is quantitative stochastic models of image degradation, their validation and calibration, and their use in characterizing the intrinsic difficulty (Bayes risk) of recognition problems and in constructing optimal classifiers and adaptive recognition systems – and, now, in precisely mapping the failure modes of machine vision systems. This is one example of the "model-directed" recognition strategy which distinguishes much of our DID area's research.

2 My Experience with HIPs

In the Fall 2000 term Prof. Richard Fateman and I taught a graduate level course "Document Image Analusis" in the CS Dept at UC Berkeley. During my discussion of document image degradation models, Prof. Manuel Blum dropped into the class and asked if these models could form the basis of a Turing Test. Right away I saw that they could, and I have been excited by the prospect ever since.

Allison Coates kindly agreed to do a class project on what we called "Pessimal Print," a means for printing text with just enough noise to baffle machines but not enough to trouble people. The heart of this project was a systematic search, helped along by a team of human volunteers and by automated tests using three excellent modern OCR machines, for ranges of parameters of our noise model where the right behavior occurs. Happily, Allison found a sweet spot: certain ranges of words, typefaces, and image degradations, described in our paper [CBF01].

Thus, e had built a CAPTCHA for all humans based on a machine-vision ability gap. The use of English words as challenges makes it, we feel, particularly comfortable for naive users of the Web including children. Our experience so far suggests that only a single challenge–word is required for very high (¿99and rejecting machines.

It is worth stressing that the gap in image pattern recognition ability between human and machine vision systems is well–known, extensively studied, and mapped quantitatively and systematically. I has been known for several years that low–quality images of printed–text documents pose particularly serious challenges to current image pattern recognition technologies [RJN96,RNN99]. In an attempt to understand the nature and severity of these challenges, models of document image degradations [Bai92,Kan96] have been developed and used to explore the limitations [HB97] of image pattern recognition

algorithms. The model of [Bai92], used in Pessimal Print, approximates ten aspects of the physics of machine–printing and imaging of text, including spatial sampling rate and error, affine spatial deformations, jitter, speckle, blurring, thresholding, and symbol size. Figure 1 shows examples of text images that were synthetically degraded according to certain parameter settings of this model.

Figure 1. Examples of synthetically generated images of machine–printed words, in various typefaces and degraded pseudo-randomly.

You – dear raeader – should be able, with little or no conscious effort, to read all these images: so will, we expect, almost every person literate in the Latin alphabet, familiar with the English language, and with some years of reading experience. The image quality of these cases of "pessimal print" is far worse than people routinely encounter, but it's not quite bad enough to defeat the human visual system.

However, many of the best present-day OCR machines are baffled by these images, as Allison Coates' data showed.

3 Current State of Pessimal Print

Kris Popat of PARC has incorporated the Pessimal Print CAPTCHA into an experimental web site which he will demonstrate at this Workshop. It is esentially Coates' method extended to dozens of typefaces. The admittedly small–scale experiments we have been able to conduct so far, on only a few thousand images, suggest that Pessimal Print has all of these highly desirable properties for a practically useful CAPTCHA:

- the test's challenges can be automatically generated;
- the number of distinct challenges generated is effectively unlimited;
- the test can be taken quickly and comfortably by human users, including children, naive Web users, and non-English speakers;
- the correct answer to each challenge is unique, unambiguous, brief, and easy to respond with;
- the test will accept virtually all human users with high reliability while rejecting very few;
- the test will reject virtually all machine users; and
- the test will (arguably) resist attack for many years even as technology advances and even if the test's algorithms are known (e.g. published and/or released as open source).

4 Future Challenges

Certainly, far larger experimental trials are needed to measure Pessimal Print's effectiveness.

Perhaps the most immediate objection to basing CAPTCHAs on machine readin is the sense, espcecially among people who have not biult OCR systems, that the problem may be easily solved — and thus the CAPTCHA defeated – with a little more engineering effort. On the other side of the argument is the fact that the pace of evolution of OCR and other species of machine vision has been slow for many decades, slow enough to provoke sharp comments in the literature [NS96,Pav00]. Machine-vision professionals with field experience with the best OCR systems are typically cautious in predicting breakthroughs. We notice that few, if any, machine vision technologies have simultaneously achieved all three of these desirable characteristics: high accuracy, full automation, and versatility. Versatility — by which we mean the ability to cope with a great variety of

types of images — is perhaps the most intractable of these, and it is the one that pessimal print, with its wide range of image quality variations, challenges most strongly.

If an arms race develops, can we (the good guys) keep ahead, given a serious effort to advance machine–vision technology, and assuming that the design principles — perhaps even the source code — of the test are known to attackers? Even given such major hints as the dictionary of words, the nature of the distortions, the fonts, sizes and other considerations, a successful attack would probably require substantially more real time than humans, at least for the near future. A statistic today suggests about 200 msec per comparison between isolated handprinted digits, using fast 2001 year workstations; many comparisons over a far larger set would be needed to solve this problem.

We can be confident that wider ranges of cases, involving other degradation parameters and other typefaces, exhibiting the right properties, can be found through straightforward experiment. Blum et al [BAL00] have experimented, on their website `www.captcha.net`, with degradations that are not only due to imperfect printing and imaging, but include color, overlapping of words, non-linear distortions, and complex or random backgrounds. The diversity of other means of bafflement ready to hand suggest to us that the range of effective text–image challenges at our disposal is enough to stay ahead of attackers.

A more difficult question to answer – and one that arises with many CAPTCHA technologies – is: how will we know if teh CAPTCHA has been broken, and so that the arms race has entered a new phase?

An ability gap exists for other species of machine vision, of course, and in the recognition of non–text images, such as line–drawings, faces, and various objects in natural scenes. One might reasonably intuit that these would be harder and so decide to use them rather than images of text. This intuition is not supported by the Cognitive Science literature on human reading of words. There is no consensus on whether recognition occurs letter–by–letter or by a word–template model [Cro82,KWB80]; some theories stress the importance of contextual clues [GKB83] from natural language and pragmatic knowledge. Furthermore, almost all research on human reading has used *perfectly formed* images of text: no theory has been proposed for mechanisms underlying the human ability to read despite extreme segmentation (merging and fragmentation) problems.

There are other, pragmatic, reasons to use images of text as challenges: the correct answer is unambiguously clear; the answer maps into a unique sequence of keystrokes; and it is straightforward automatically to label every challenge, even among hundreds of millions of distinct ones, with its answer. These advantages are lacking, or harder to achieve, for images of objects or natural scenes.

It might be good in the future to locate the limits of human reading in our degradation space: that is, at what point do humans find degraded words unreadable; do we smoothly decay or do we show the same kind of "falling off a cliff" phenomenon as machines but just at another level?

5 Acknowledgments

The project was triggered by a question – could character images make a good Turing test? – raised in the class by Manuel Blum of Carnegie–Mellon Univ., as one possible solution of the "chat room problem" posed earlier by Udi Manber of Yahoo!. Manuel Blum, Luis A. von Ahn, and John Langford, all of CMU, shared with us much of their early thinking about automated Turing tests, including news of their CAPTCHA project and their website `www.captcha.net`, which influenced the design and evolution of our Pessimal Print project. The Pessimal Print web site was built by Kris Popat. I wish gratefully to acknowledge stimulating discussions with Bela Julesz in the late 1980's (at Bell Labs) on the feasibility of designing "pessimal fonts" using textons, which were however never implemented.

6 Bibliography

[CBF01] A. L. Coates, H. S. Baird, and R. J. Fateman, "Pessimal Print: a Reverse Turing Test," *Proc., IAPR 6th Int'l Conf. on Doc. Anal. and Recogn*, Seattle, WA, September 10-13, 2001, pp. 1154-1158.

[BAL00] M. Blum, L. A. von Ahn, and J. Langford, *the CAPTCHA Project*, "Completely Automatic Public Turing Test to tell Computers and Humans Apart," `www.captcha.net`, Dept. of Computer Science, Carnegie–Mellon Univ., and personal communications, November, 2000.

[Bai92] H. S. Baird, "Document Image Defect Models," in H. S. Baird, H. Bunke, and K. Yamamoto (Eds.), *Structured Document Image Analysis*, Springer–Verlag: New York, 1992, pp. 546-556.

[Cro82] R.G. Crowder, *The Psychology of Reading*, Oxford University Press, 1982.

[GKB83] L. M. Gentile, M. L. Kamil, J. S. Blanchard *'Reading Research Revisited* , Charles E. Merrill Publishing, 1983.

[HB97] T. K. Ho and H. S. Baird, "Large-Scale Simulation Studies in Image Pattern Recognition," *IEEE Trans. on PAMI*, Vol. 19, No. 10, pp. 1067–1079, October 1997.

[ISRI] Information Science Research Institute, University of Nevada, Las Vegas, 4505 Maryland Parkway, Box 454021, Las Vegas, Vevada 89154-4021 USA.

[Jen93] F. Jenkins, *The Use of Synthesized Images to Evaluate the Performance of OCR Devices and Algorithms*, Master's Thesis, University of Nevada, Las Vegas, August, 1993.

[Kan96] T. Kanungo, *Document Degradation Models and Methodology for Degradation Model Validation*, Ph.D. Dissertation, Dept. EE, Univ. Washington, March 1996.

[KWB80] P.A. Kolers, M. E. Wrolstad, H. Bouma, *Processing of Visible Language 2*, Plenum Press, 1980.

[NS96] G. Nagy and Seth, "Modern optical character recognition." in *The Froehlich/Kent Encyclopaedia of Telecommunications*, Vol. 11, pp. 473-531, Marcel Dekker, NY 1996.

[Pav00] T. Pavlidis, "Thirty Years at the Pattern Recognition Front," King-Sun Fu Prize Lecture, 11th ICPR, Barcelona, September, 2000.

[RNN99] S. V. Rice, G. Nagy, and T. A. Nartker, *OCR: An Illustrated Guide to the Frontier*, Kluwer Academic Publishers, 1999.

[RJN96] S. V. Rice, F. R. Jenkins, and T. A. Nartker, "The Fifth Annual Test of OCR Accuracy," ISRI TR-96-01, Univ. of Nevada, Las Vegas, 1996.

[SCA00] A. P. Saygin, I. Cicekli, and V. Akman, "Turing Test: 50 Years Later," *Minds and Machines*, 10(4), Kluwer, 2000..

[Spi97] A. L. Spitz, "Moby Dick meets GEOCR: Lexical Considerations in Word Recognition," *Proc., 4th Int'l Conf. on Document Analysis & Recog'n.*, Ulm, Germany, pp. 221–232, August 18–20, 1997.

[TS81] O. J. L. Tzeng and H. Singer, *Perception of Print: Reading Research in Experimental Psychology*, Lawrence Erlbaum Associates, Inc., 1981.

[Tur50] A. Turing, "Computing Machinery and Intelligence," *Mind*, Vol. 59(236), pp. 433–460, 1950.