The Influence of Image Complexity on Handwriting Recognition

Amalia Rusu, Venu Govindaraju

The Influence of Image Complexity on Handwriting Recognition

Amalia Rusu and Venu Govindaraju
University at Buffalo, The State University of New York
Buffalo, NY, USA
{air2,govind}@cedar.buffalo.edu

Abstract

Automated recognition of unconstrained handwriting continues to be a challenging research task. In addition to the errors caused by image quality, image features, segmentation, and recognition, in this paper we have also explored the influence of image complexity on handwriting recognition and compared humans' versus machines' recognition. We describe a new methodology that will exploit the gap between the abilities of humans and computers in reading handwritten text images and investigate the influence of handwritten image complexity and Gestalt Laws of perception on this gap. Experimental results are presented and compared for image density and perimetric complexity of handwritten challenges. We make use of current challenges in handwriting recognition for applications in Cyber security.

Keywords: handwriting recognition, image complexity, CAPTCHA.

1. Introduction

Our paper is focused on exploring the gap between human and machine abilities in reading handwriting under noisy conditions and determining a set of transformations that alter various aspects of handwriting such as legibility, overlapping of words, broken strokes, and the extent of overrun characters. Furthermore we explore the relationship between several types of image transformations and image complexity, and determine how this affects the ability of humans and machines to perform recognition. Whereas the ultimate objective of Artificial Intelligence is to build machines which can demonstrate human-level abilities, we explore the limitations of machines in handwriting recognition tasks and identify new applications where these same limitations are actually an advantage. Consequently these limitations can help solve problems in computer security.

In general the successful handwriting applications are all characterized by small or fixed lexicons afforded by contextual knowledge. However, recognition of unconstrained handwriting is difficult because of the diversity in writing styles, inconsistent spacing between words and lines, and uncertainty of the number of lines in a page as well as the number of words in a line, and prior knowledge of the context. In addition, current handwritten word recognition approaches depend on the availability of a lexicon of words for matching, making the recognition accuracy dependent upon the size of the lexicon. So, for a truly general application-independent word recognizer, the lexicon would be a large subset of the English dictionary and the accuracy of recognition would be very low.

In contrast to low machine recognition accuracy, our experiments reaffirm the superiority of humans in reading handwritten text especially under conditions of low image quality, clutter and occlusion. In our attempt to quantify the strengths of human reading abilities we have encountered controversial results. In our current experiments neither image density nor perimetric complexity has been shown to predict the efficiency of humans in handwritten word recognition. On the other hand Gestalt and Geon components have been shown to play an important role in the relative identification of characters and at the same time pose problems to machines' recognition.

2. Handwritten Challenges Generation

We use a large repository of handwritten word images that current handwriting recognizers cannot read if small deformations are applied, even when given a lexicon. Examples of original handwritten city-word images are shown in Figure 1.

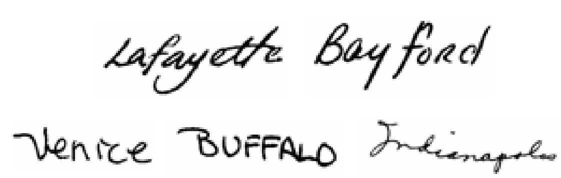

Figure 1. Examples of original handwritten US city name images.

Automatic generation of synthetic handwritten samples using models such as the ones described in [11], [12] will be a valid option for creating infinite-many distinct challenges. The goal is to design handwritten word images that exploit the knowledge of the common source of errors in automated handwriting recognition systems and at the same time take advantage of the salient aspects of human reading. For example, humans can tolerate intermittent breaks in strokes (using the Gestalt law of continuity) but current programs fail when the breaks vary in size or exceed certain thresholds. The simultaneous interplay of several Gestalt laws of perception adds to the challenge of finding the range of parameters that separate human and machine abilities.

We are interested in analyzing the holistic aspects of human reading. Gestalt psychology is based on the observation that we often experience things that are not part of our simple sensations [6]. What we see is believed to be an effect of the whole event, which is more than the sum of the parts. This concept is similar to the holistic word recognition approaches that focus on recognizing the entire word at once [7]. The Gestalt laws of organization include: proximity (grouping by distance or location), similarity (grouping by type), symmetry (grouping by meaning), continuity (grouping by flow of lines or by alignment), closure (perceive shapes that are not completely there), familiarity (elements are more likely to form groups if the groups appear familiar), and figure-ground distinction (see Figure 2). Although memory is not perception-based, it plays a role in perception as an outside iconic memory with internal metric relations.

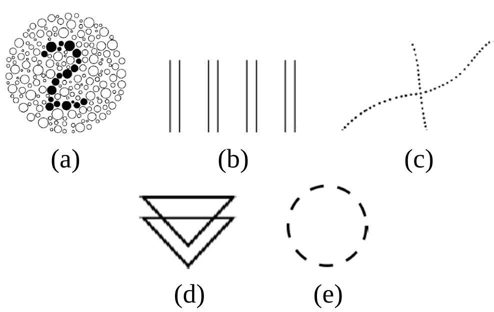

Figure 2. Several examples for Gestalt Laws of perception: (a) similarity, (b) proximity, (c) continuity, (d) symmetry, (e) closure.

Segmentation is a complex and computationally expensive step which usually creates problems to recognition if not effectively solved. Our task was to remove features or add non-textual strokes or noise to a handwritten image in a systematic fashion based on Gestalt segmentation and grouping principles in order to defeat machine recognition, while preserving the overall letter legibility and word recognition for humans. We have also noticed that Geon theory [3] of pattern recognition, which refers to recognition by components, provides good hints on what is desirable to be preserved for an easy image reconstruction. For example Figure 3 suggests that preserving edges and intersections helps humans recover the objects, and in particular it is true for handwritten characters where junctions, crossing strokes, concavities and convexities overlaps are important features as well.

We started by applying Gestalt laws on handwritten strokes. We have found that the laws can be translated into methods that can be controlled by adjusting parameters and used as transformations on handwritten images. We consider several sets of candidate transforms that correspond to these laws and control them for readability.

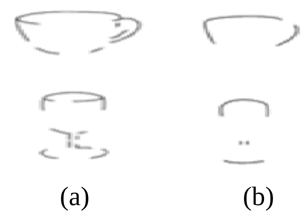

Figure 3. Evidence of Geon Theory. (a) Recoverable objects, (b) Non-recoverable objects.

We considered the following transformations in our tests: creating horizontal or vertical overlaps, adding occlusions (i.e. circles, rectangles, lines with random angles), adding occlusions by waves from left to right on entire image, with various amplitudes and wavelength or rotate them by an angle, adding occlusion using the same pixels as the foreground pixels, arcs, or lines, with various thicknesses (in general, curved strokes could be confused with part of a character and using asymmetric strokes prevent against learning the pattern), using empty letters, broken letters, edgy contour, fragmentation (break characters such that general image processing techniques cannot reconstruct the original image), splitting the image in two parts on horizontal and displace the parts in opposite directions, splitting the images in parts, either by a vertical/horizontal line or by diagonals, and spread the parts apart (i.e. mosaic effect) - symmetry in displacement helps image reconstruction for humans, or adding occlusion using the same pixels as the foreground pixels, or black jaws (extra strokes are confused with character components when they are about the same size, thickness, and curvature as the handwritten characters). Several examples of handwritten challenges obtained through these transformations can be seen in Figure 4.

Figure 4. Examples of transformations that defeat current OCR systems. The truth words are: *Victor, Rochester, Buffalo, Lockport, Niagara, Scottsville, West, Pittsburgh*.

3. Computation of Image Complexity

We investigated the image complexity of our handwritten challenges and explored several approaches to determine it. Since human visual perception is sensitive to contrast and border perception we used the perimetric complexity and image density as a factor. We considered image density as the number of black pixels versus the total number of pixels in an image. We also measured perimetric complexity, defined as the inside and outside perimeter squared divided by black area (Eq. 1) and used the algorithm described in [8] to compute it.

$$PerimetricComplexity = P^2/A \quad (1)$$

Since we work with binary images, the black area is the number of 1's. To compute the perimeter we first generated the image contour 1-pixel thick and then thicken it to obtain a 3-pixel thick image and divide the perimeter to 3 so that we count diagonal as height plus weight. For stroked characters, it is known in literature that the perimetric complexity captures how convoluted a character is, and it is also easily computed, independent of size, additive, as well as suitable for binary images. On the other hand perimetric complexity might help explain handwritten strokes identification. The reverse of perimetric complexity is also used in pattern recognition and known as object compactness (Eq. 2).

$$Compactness = A/P^2 \quad (2)$$

In addition, we have explored the relationship between human efficiency in reading the handwritten challenges and image complexity for our handwritten samples as well as the effect of each transformation that we considered on the image complexity and indirectly on machines' recognition.

4. Word Recognizers and Results

We have generated test images to be recognized by state-of-the-art handwriting recognizers: Word Model Recognizer (WMR), Character Model Recognizer (CMR), and Accuscript [4], [5], [13]. WMR is a segmentation-based recognizer that treats each word as a model and finds 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. Accuscript is a grapheme-based recognizer, which extracts features from sub-characters such as loops, turns, junctions, arcs, without explicit segmentation. CMR follows a segmentation-then-recognition scheme. The algorithm consists in modules for image preprocessing, segmentation, character recognition and lexicon ranking. The goal is to segment, isolate and recognize the characters which make up the word. All recognizers described take the advantage of using static lexicons in the recognition process as well as using preprocessing techniques to enhance image quality and remove noise, correct the slant, and smooth the contour, thus making the performance evaluation of machines a fair test.

Our experiments include the methods described in association with the Gestalt laws of perception. Several new sets of transformed images were used - one set for each deformation method previously described. We randomly choose parameter values for our transformations and apply them on handwritten word images. We assumed that a valid lexicon is provided and that for every image the corresponding truth word was always present in the lexicon. We ran tests on lexicons of size about 4000 and 40000 - corresponding to the entire list of US city names, for the case of a context tide to city names only.

We also used random subsets of about 90 images each to be recognized by voluntary students. The test consists of about 10 handwritten word images per each type of transformation previously described. The images were chosen at random for each student from the sets of about 4100 deformed images per transformation that have been tested on the recognizers as well. The human subjects were relatively familiar with the words in the images since they are city names in the U.S. The tests suggest that human performance depends on context, and prior knowledge of the word provides the greatest advantage and can influence human perception. This relates to gestalt principles of memory and word familiarity which could be used as an advantage for humans. The human results are presented in Table 1 and the accuracies achieved by our testing recognizers are presented in Table 2. The transformations are as follows: 1. All Transformations, 2. Empty Letters, 3. Small Fragmentation, 4. High Fragmentation, 5. Displacement, 6. Mosaic, 7. Jaws/Arcs, 8. Occlusion by circles, 9. Occlusion by waves, 10. Black Waves, 11. Vertical Overlap, 12. Horizontal Overlap Small, 13. Horizontal Overlap Large. The gap in the ability in recognizing handwritten text between humans and computers is illustrated in Figure 5. We considered an answer to be correct when all the characters of a handwritten word challenge had been recognized.

Table 1. The accuracy of human readers for a set of randomly chosen 1069 images. Most of the human errors came from junk original images rather than difficulties with the deformations applied on those images. Usually if the original handwritten sample was clean to begin with, after deformation did not create problems for humans to recognize the word.

Transformations	*Number of Tested Images*	*Accuracy*
1.	1069	76.1
2.	89	82.0
3.	88	73.9
4.	90	74.4
5.	89	78.6
6.	90	74.4
7.	89	71.9
8.	90	67.8
9.	87	80.5
10.	90	80.0
11.	88	87.5
12.	90	76.7
13.	89	65.2

5. Experiments on Image Complexity

We computed image density and perimetric complexity for all generated handwritten challenges. As Figure 6

Table 2. The accuracy of handwriting recognizers (HR). A set of 4100 images was tested for each kind of transformation using lexicons with size 4000 and 40000.

HR	*WMR*		*Accuscript*		*CMR*	
	4000	*40000*	*4000*	*40000*	*4000*	*40000*
1.	14.4	6.6	7.0	2.9	5.2	3.8
2.	0.9	0.4	0.1	0.0	1.2	0.9
3.	0.0	0.0	0.2	0.2	0.0	0.0
4.	0.5	0.2	0.0	0.0	0.3	0.2
5.	19.7	10.3	8.8	3.4	3.5	2.4
6.	14.3	6.4	9.0	3.0	3.1	2.2
7.	5.1	1.3	3.6	0.8	1.7	1.3
8.	35.9	20.3	32.3	17.4	25.6	19.9
9.	15.4	7.0	10.6	4.3	5.8	4.2
10.	16.4	5.3	1.6	0.4	5.7	4.3
11.	27.9	14.4	12.6	3.9	6.8	5.1
12.	24.3	10.7	2.9	0.6	4.9	3.2
13.	12.9	3.6	2.4	0.4	3.4	2.1

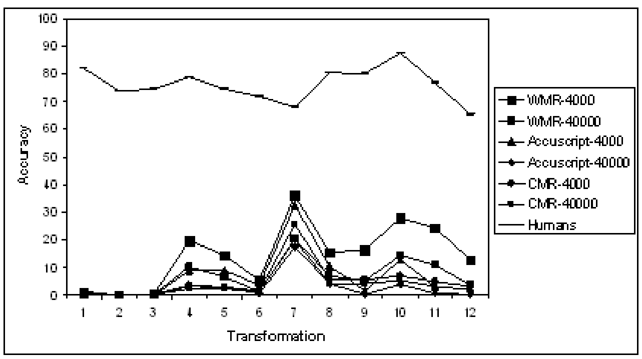

Figure 5. Ability gap in recognizing handwritten text between humans and computers per type of transformation (empty letters, fragmentation small/high, displacement, mosaic, jaws, occlusion by circles, occlusion by waves, waves, vertical overlap, horizontal overlap small and large).

shows, several transformations could be used if we want to increase challenges' complexity. By using empty letters and fragmentations, the image complexity drastically increases, whereas for the rest of transformations the image complexity is relatively comparable (i.e. adding arcs and jaws, mosaic effect, overlaps). A first observation in our study is that the transformations corresponding to empty letters and fragmentations have the smallest machine recognition accuracy, as shown in Table 2, but on the other hand the generated challenges have the highest image complexity. Based on this observation we can assume that using handwritten images with higher image perimetric complexity is an alternative way of lowering machine's accuracy. We have also generated images with same image density but different perimetric complexity, and conclude that image density is not a good indicator for the complexity of the challenges in our case and does not correlate well with human recognition efficiency (see Figure 7). The snapshots in Figure 6 and Figure 7 are for the same 20 sample images. We observe from the charts that in these cases low image density corresponds to high image complexity and vice-versa which is intuitively correct giving the computational formulas.

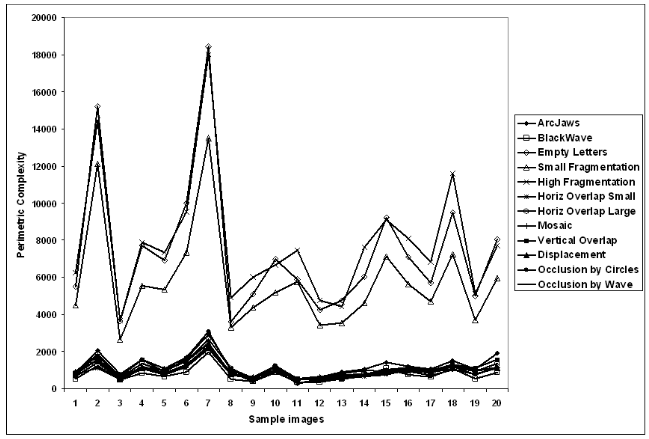

Figure 6. A snapshot of handwritten images and corresponding perimetric complexity (perimeter squared divided by area of black pixels).

Unlike the results reported by [8], in our experiments perimetric complexity does not correlate well with human recognition accuracy for 1058 distinct sample images tested on human subjects (see Figure 8). However, these results support the importance of other factors involved in human handwriting recognition such as the role of Gestalt principles, as well as preservation of Geon components such as intersections and edges, and having prior knowledge of the context. Similar non-correlations between the perimetric complexity and humans' recognition efficiency have been reported in [1] for machine printed word images.

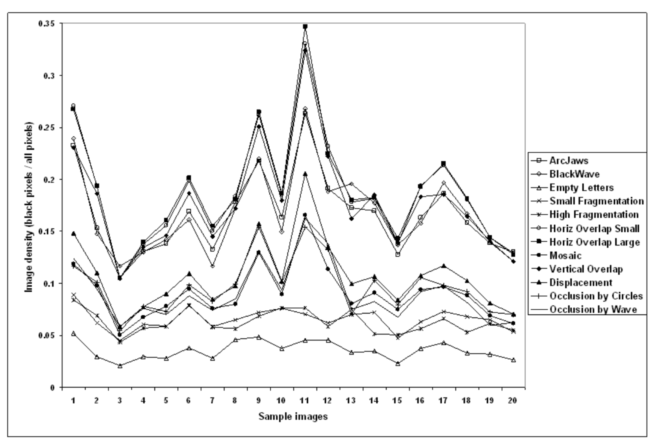

Figure 7. A snapshot of handwritten images and corresponding image density (black pixels vs. total pixels).

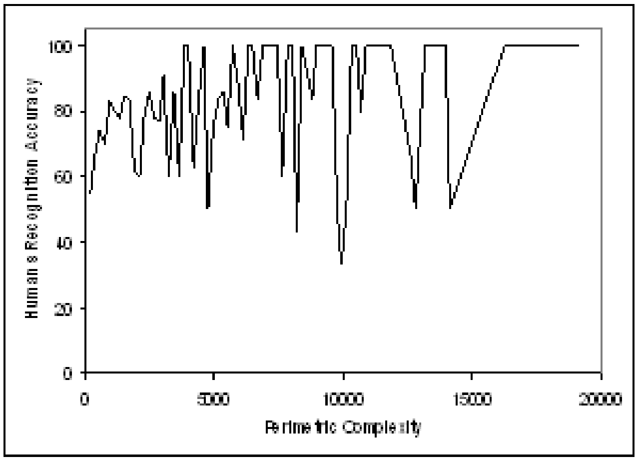

Figure 8. Humans recognition accuracy vs. perimetric complexity as a percent of correct answers per bin (with a total range for perimetric complexity of 100 equal bins, complexity range 0-20000).

6. Summary and Future Work

We have identified the gap between humans and computers in reading handwritten text images and propose to use it to defend Web services against bot attacks. In our previous work, we already established a new application of handwriting recognition to CAPTCHAs - *Completely Automatic Public Turing test to tell Computers and Humans Apart*, which exploits the difference in the reading proficiency between humans and computers when dealing with handwritten text images [9]. In the context of Alan Turing's theory, where he proposed ways of testing whether machines can think through experiments that involve interrogation of both human and computer and where the interrogator should be able to distinguish between them, these new tests are considered reverse Turing tests [2], [10]. Since CAPTCHAs exploit the areas where computers are not as good as humans (yet), handwriting recognition is a strong candidate for these tests.

We make use of current challenges in handwriting recognition for applications in Cyber security. Our handwritten challenges can help solve problems in computer security that other existing security measures cannot address properly. An important part of designing a reliable system with adjustable challenges is the parameterization of the level of difficulty of our handwritten challenges. We considered several options to determine the image complexity of our challenges and explored their influence on handwriting recognition. We have also started to build a website at *http://www.cedar.buffalo.edu/~air2/captcha/captcha.php* and invite all interested human users and their computer programs to attack our handwritten challenges and provide feedback to the authors. Experimental results on three handwriting recognizers have shown the gap in the ability between humans and computers in handwriting recognition. We also conducted preliminary user studies and human surveys on handwritten challenges and the analysis of results strongly supports our hypothesis. However, we have found that human efficiency does not correlate well with image complexity of handwritten word challenges described by either image density or perimetric complexity, leaving the reason of successful human recognition on other heuristic Gestalt assumptions. Generation of synthetic handwritten samples will be investigated such that we could test the influence of any real or non-sense words on humans' and machines' recognition and further attempt to quantify the gap between them. We also raised some questions on recognizing transformed images and further need to investigate that. If becoming a research interest, computer programs may improve in that domain to reduce the gap between humans' and machines' recognition and increase recognizers' robustness.

References

[1] H. S. Baird, M. A. Moll, S. Wang "ScatterType: A Legible but Hard-to-Segment CAPTCHA", *Proceedings of the Eighth International Conference on Document Analysis and Recognition*, 2005, pp 935-939.

[2] H. Baird, and K. Popat, "Human interactive proofs and document image analysis", *Proceedings of the IAPR Workshop on Document Analysis Systems*, 2002.

[3] I. Biederman, "Higher-Level Vision", *Visual Cognition and Action*, 1990.

[4] J. T. Favata, "Character model word recognition", *Proceedings of the Fifth International Workshop on Frontiers in Handwriting Recognition*, 1996, pp 437-440.

[5] 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*, 1997, pp 366-379.

[6] K. Koffka, *Principles of Gestalt Psychology*, 1935.

[7] S. Madhvanath, and V. Govindaraju, "The role of holistic paradigms in handwritten word recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001.

[8] D. G. Pelli, C. W. Burns, B. Farell, and D. C. Moore, "Identifying letters", *Vision Research*, 2005.

[9] A. Rusu, and V. Govindaraju, "Handwritten CAPTCHA: Using the difference in the abilities of humans and machines to read handwritten words", *Proceedings of the Ninth IAPR International Workshop on Frontiers of Handwriting Recognition*, 2004, pp 226-231.

[10] A. Turing, "Computing machinery and intelligence", *Mind*, 1950, pp 433-460.

[11] T. Varga, and H. Bunke, "Offline Handwriting Recognition Using Synthetic Training Data Produced By Means Of A Geometrical Distortion Model", *IJPRAI*, 2004, pp 1285-1302.

[12] T. Varga, D. Kilchhofer, and H. Bunke, "Template-based Synthetic Handwriting Generation for the Training of Recognition Systems", *Proceedings of the 12th Conference of the International Graphonomics Society*, 2005, pp 206-211.

[13] H. Xue, and V. Govindaraju, "A stochastic model combining discrete symbols and continuous attributes and its applications to handwriting recognition", *Proceeding of the International Workshop on Document Analysis and Systems*, 2002, pp 70-81.