9-1-2007

Human Image Preference and Document Degradation Models

Chris Hale
Boise State University

Elisa H. Barney Smith
Boise State University

DOI: 10.1109/ICDAR.2007.4378715

Human Image Preference and Document Degradation Models

Chris Hale and Elisa H. Barney Smith
Electrical and Computer Engineering, Boise State University
Boise, ID 83725-2075 USA
EBarneySmith@boisestate.edu

Abstract

Because most degraded documents are created by people, the preferences individuals have in relation to degraded documents are quite important. Their preferences may determine whether or not the documents they created are appropriate for machines. The goal of this study was to find relationships between preference and several parameters of a scanner degradation model. It was found that the difference in binarization threshold and the difference in edge displacement caused by the degradation both had strong linear relationships to preference. The width of the point spread function did not show such a relationship. These relationships were counterintuitive because degraded characters with thicker stroke widths than the original were preferred to those that had stroke widths closer to the original character.

1. Introduction

When documents are digitized on a desk top scanner, the image is degraded. For the same resolution, the level of degradation is greater, or the Optical Character Recognition (OCR) accuracy is lower [8, 12], if only a thresholded binary image is retained. The effect of a poorly digitized document image on OCR has been documented [13], but the definition of poorly digitized is reliant on the performance of the OCR engine for those studies. Therefore, the definition is circular.

The issues with digitization quality in large digital library projects was explored by a working group at the DIAL 04 workshop [5]. They encouraged test charts to be used to determine the best scanner setting and to let the calibration be automated to provide the best input image for OCR. In reality, the acquisition and decisions about its quality control will be done by humans, usually untrained in acquisition quality. They will not intentionally make a poor quality document, but it still often happens. In fact, there are differences between naïve users and experts when judging the quality of black and white images [9]. This paper investigates the correlation between the perceived image quality as perceived by non document specialists and quantifiable degradations based on a mathematical degradation model.

If human preferences coincide with the categorization that lead to improved OCR performance, untrained human operators may make good decisions about how to acquire an image for input into an OCR package. However, if the preferences follow another trend, untrained operators would most likely make bad decisions about OCR input. Also, existing bilevel scans could follow preferences because it is likely the human operators would use their own judgment to choose the scanning parameters. Therefore, the preferences could provide insight into the values of the degradation parameters for previously scanned and thresholded documents.

2. Degradation Model

The degradation model used for this research is based on the model developed by Baird [1] where a bilevel image is blurred through two dimensional convolution with a Point Spread Function (PSF) with width w, then thresholded at a binarization threshold Θ to produce a bilevel image. The model contains other components, but these are the two most significant parameters in the model affecting degradations of bilevel images [10]. In this work, the PSF is assumed to be a bivariate Gaussian with the width, w, equal to the standard deviation measured in units of pixels. The threshold is measured in absorptance in a range [0,1].

The raw model parameters, w and Θ have a combined effect on the characters in Figure 1. Two resulting effects of the degradation have been defined: the amount an edge is displaced, δ_c, and the amount a black or white corner is eroded [3, 4]. A statistical test was conducted in [2] to compare the similarity between groups of characters synthetically generated with parameters (w, Θ) varying over the parameter space. This test showed that the amount of variation in the characters correlated highly with the amount of edge spread,

$$\delta_c = -w\, \mathrm{ESF}^{-1}(\Theta)\,, \qquad (1)$$

where ESF( ) is the Edge Spread Function, which is the integral of the PSF. Character images produced using sets of parameters with a common δ_c did not produce as large a

Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)
0-7695-2822-8/07 $25.00 © 2007 IEEE

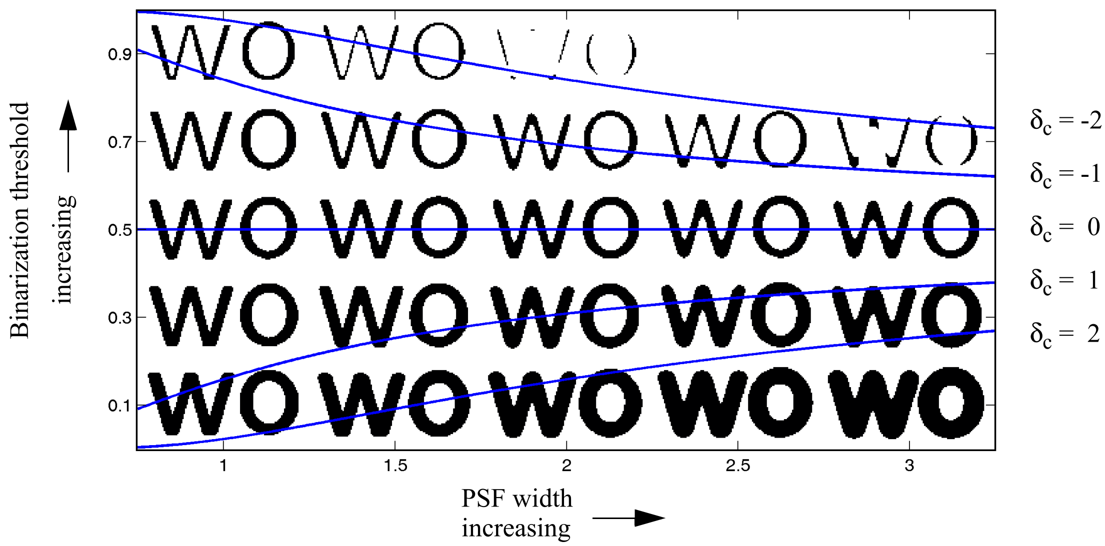

Figure 1: Characters after blurring and thresholding over a range of PSF widths, *w*, and binarization thresholds, Θ. A broad range of character appearances can be seen, but certain characters have some general similarities.

difference between characters as those between pairs of characters with other degradation model parameters.

Other studies have shown that using edge spread curves to partition the degradation space can improve OCR performance by grouping characters that are similar based on the features used for classification [6, 7]. This further solidifies the idea that the edge spread curves are a natural way to categorize items in the degradation space. The work in this paper explores whether human preferences also follow the edge spread categorization, or even the PSF width or binarization threshold (*w* or Θ) directly.

3. Experiments and Data

A survey was conducted to compare pairs of characters, each degraded with different model parameters. The letters shown to participants were sans-serif characters, ‘w’ and ‘o.’ These were chosen so there would be a character with many sharp corners and a character with no corners. These were also the characters used in [2]. The survey used samples with five different edge spread amounts and three different PSF widths for a total of 15 combinations of parameters for each character. Figure 1 shows the characters w and o each degraded over a variety of PSF widths and binarization thresholds without noise. Superimposed on this are lines of equal edge displacement at δ_c={-2, -1, 0, 1, 2}. The characters used in the study were created so they would be along one of these lines. The degradations were compared pair-wise with every other degradation, for a total of 120 comparisons.

The survey was administered on a computer showing in turn images of the two degraded characters with a slider below them. The images were reversed from comparison to comparison to prevent a left/right bias. The participants chose which image in each pair looked better by moving the slider position toward the image they preferred. The distance the slider was moved corresponded to how much better the preferred image looked. The result of the comparison was a number between negative one and one as determined by the slider’s position. This method is essentially an electronic version of the graphic rating scale [11], also called the visual analogue scale used often in psychophysical experiments. In this case, the position between the two images on the scale is chosen with the slider instead of the more traditional pencil and paper. The electronic version used in this experiment allows for much simpler scoring than the traditional version with a minimal loss in resolution.

The survey was administered to 93 participants. The participants were volunteers from an introductory psychology course so they had no extensive background in document degradation analysis. The participants were told to pick the letter that looks better. No other selection criteria were given.

4. Results

The raw data consisted of the edge spread, the PSF width, and the preference score for each character displayed. The threshold values were calculated using Equation 1. The comparisons were classified by the difference in width value, *w*, threshold value, Θ, and edge spread value, δ_c between the characters being compared.

Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)
0-7695-2822-8/07 $25.00 © 2007 IEEE

Table 1: Correlation data using width difference for 'w', with means shown on the diagonal.

Width Difference	0	1	2
0	(-0.37)	-0.39	-0.38
1	-0.39	(0.00)	0.62
2	-0.38	0.62	(-0.06)

Table 2: Correlation data using width difference for 'o' with means shown on the diagonal.

Width Difference	0	1	2
0	(-0.20)	-0.26	-0.43
1	-0.26	(-0.03)	0.62
2	-0.43	0.62	(-0.09)

Table 3: Correlation data using threshold difference for 'w' with means shown on the diagonal

Threshold Difference	0	(0,0.4)	(0.4,1)
0	(-0.03)	ns	ns
(0,0.4)	ns	(-0.36)	0.77
(0.4,1)	ns	0.77	(-0.71)

Table 4: Correlation data using threshold difference for 'o' with means shown on the diagonal

Threshold Difference	0	(0,0.4)	(0.4,1)
0	(-0.05)	ns	ns
(0,0.4)	ns	(-0.16)	0.93
(0.4,1)	ns	0.93	(-0.37)

After acquiring the raw data, the data for each participant was grouped by averaging the responses for all image pairs where the PSF width difference was common, the edge spread difference was common or the threshold difference was within certain ranges. This resulted in three width differences, five threshold differences, and nine edge spread differences each with 93 samples. These items were used in all of the data analysis.

The correlation between the different values was calculated. The results of the correlations of the variables from the condensed sets are shown in Tables 1-6. The mean value of each variable is shown in the main diagonal of the correlation matrices. Correlations that are not significant, with $p < .05$, are indicated with 'ns.' The correlation's significance shows how consistent the participants were in evaluating pairs of comparisons. The magnitude of the

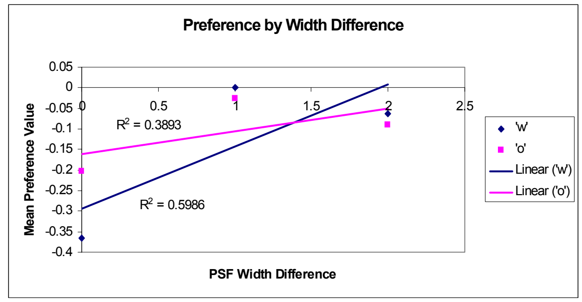

Figure 2: Mean preference value using PSF width difference

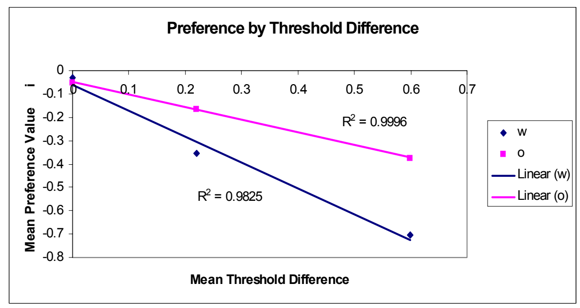

Figure 3: Mean preference value using threshold difference

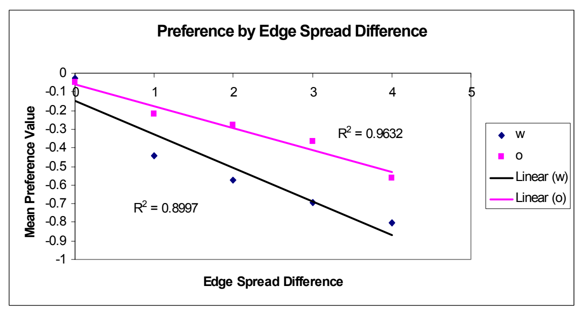

Figure 4: Mean preference value using edge spread difference.

value shows the strength of relationship of the two variables.

Tables 1 and 2 show the correlations and means for comparisons by the difference in PSF width. The correlations indicate that the participants reported similar preferences. However, the mean values were mostly near zero, so there was little difference in preference for different widths.

Looking at how the threshold difference affected user preference yielded mostly significant correlations, Tables 3 and 4. The only correlations that were not significant involved a threshold difference of zero. Therefore, the participants had similar preferences in every case except for the items with the same threshold values. The means

0-7695-2822-8/07 $25.00 © 2007 IEEE

Table 5: Correlation data using edge spread difference for 'w' with means shown on the diagonal

Edge Spread Difference	0	1	2	3	4
0	(-0.03)	ns	ns	ns	0.28
1	ns	(-0.44)	0.90	0.76	0.48
2	ns	0.90	(-0.58)	0.90	0.66
3	ns	0.76	0.90	(-0.70)	0.83
4	0.28	0.48	0.66	0.83	(-0.80)

Table 6: Correlation data using edge spread difference for 'o' with means shown on the diagonal

Edge Spread Difference	0	1	2	3	4
0	(-0.05)	ns	ns	ns	ns
1	ns	(-0.22)	0.95	0.90	0.73
2	ns	0.95	(-0.28)	0.95	0.76
3	ns	0.90	0.95	(-0.36)	0.83
4	ns	0.73	0.76	0.83	(-0.56)

also indicate that lower threshold values were preferred to higher threshold values.

Most of the correlations using the edge spread difference were also significant, Tables 5 and 6. The strength of the significant correlations increases as the values of the difference in edge spreads get closer. The only correlations that were not significant involved one comparison where the images had the same edge spread. It is possible that, with a mean near zero, the results for the comparisons of equal edge spreads were not consistent enough for a significant relationship.

To determine if there was a clear trend in the preferences, a series of linear regressions were run on the data. The regressions show how good of a linear relationship exists among the mean preference values for each category. The mean preference values were the means of all the participants' preference for each value of a category. The values for the threshold category were the means of each instance in the Θ ranges.

Figure 2 shows that the linear regression using the PSF width data did not reveal a good linear relationship. However, the threshold data and the edge spread data did reveal a good linear relationship for each letter, Figures 3 and 4. These findings imply that there is a stronger relationship between threshold and preference than between PSF width and preference. The edge spread also has a stronger relationship than PSF width. In addition, the edge spreads that result in characters with thicker strokes were preferred to those producing thinner ones, and the lower threshold value was preferred, which results in a thicker stroke. As a result, characters that have a thicker stroke width than the original are preferred to those that have the same stroke width as the original image.

5. Conclusions and Future Work

There is a strong indication that the threshold and edge spread play a significant role in personal preference. The relationship is unlikely to be as linear as indicated in Figures 3 and 4. The study limited the degradation space to characters that would be legible. However, if a larger degradation space had been used, illegible characters should make the relationship less linear. It is still possible that the relationship may be piece-wise linear which would increase its usefulness for modeling purposes.

Although it appeared that the relationship between PSF width and preference was weak, there may have been an insufficient sample of the degradation space. It is possible that one of the three points is an outlier and the remaining points with more width samples form a better linear relationship. Therefore, it will be important to test more PSF width values in future studies. A greater PSF width range will also introduce illegible characters.

Another important consideration is the fact that this model did not include noise. While noise is associated with characters having a poor appearance, noise was omitted because the time allowed to administer the study to the participants was insufficient to allow for enough comparisons to include noise. Noise should be more apparent in characters with a greater edge spread. So, the addition of noise could make the characters with a stroke width closer to that of the original more appealing.

Furthermore, the thicker stroke widths were preferred to the thinner. However, the characters were shown in isola-

Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)
0-7695-2822-8/07 $25.00 © 2007 IEEE

tion. If the characters were displayed in a string or text was used, the character spacing may affect the results either by the context of the other characters in a string or by causing characters to touch when their strokes are thickened. Even when the degradation caused corner erosion, the thicker characters were still preferred. Additional research should be conducted to determine whether the degraded characters would be preferred to bolded, ideal characters with the same stroke width.

When the additional research is completed, the results should provide insight into the ability of untrained operators to make reasonable choices while digitizing text for OCR. The results could provide information regarding which threshold is ideal for human viewing based on the scanner parameters.

6. Acknowledgement

This material is based on work supported by the National Science Foundation under grant No. CCR-0238285.

7. References

[1] Henry S. Baird, "Document Image Defect Models," *Proc. IAPR Workshop on Syntactic and Structural Pattern Recognition*, Murry Hill, NJ, June 1990, pp. 13-15. Reprinted in H. S. Baird, H. Bunke, and K. Yamamoto (Eds.), *Structured Document Image Analysis*, Springer Verlag: New York, 1992, pp. 546-556.

[2] Elisa H. Barney Smith and Xiaohui Qiu, "Statistical image differences, degradation features and character distance metrics," International Journal of Document Analysis and Recognition, Vol.6, No. 3, 2004, pp. 146-153.

[3] Elisa H. Barney Smith, "Characterization of Image Degradation Caused by Scanning," *Pattern Recognition Letters*, Vol. 19, No. 13, 1998, pp. 1191-1197.

[4] Elisa H. Barney Smith, "Estimating Scanning Characteristics from Corners in Bilevel Images," Proc. SPIE Document Recognition and Retrieval VIII, Vol. 4307, San Jose, CA, 21-26 January 2001, pp.176-183.

[5] Elisa H. Barney Smith, Henry Baird, William Barrett, Frank Le Bourgeois, Xiaofan Lin, George Nagy and Steve Simske, "DIAL 2004 Working Group Report on Acquisition Quality Control," Proc. 2nd International Workshop on Document Image Analysis for Libraries 2006, Lyon, France, April 2006, pp. 373-376.

[6] Elisa H. Barney Smith and Tim Andersen, "Partitioning of the degradation space for OCR training," Proc. SPIE Electronic Imaging, Document Recognition and Retrieval XIII, Vol. 6067, San Jose, CA, January 2006, paper #606705.

[7] Elisa H. Barney Smith and Tim Andersen, "Text Degradations and OCR Training," International Conference on Document Analysis and Recognition 2005, Seoul, Korea, 29 August - 1 September 2005, pp. 834-838.

[8] William Barrett and Rex Barzee, "Posting paper on the web," Proc. Vision Interface '98, Vancouver, BC, Canada, 1998, pp. 381-388.

[9] Chengwu Cui, "Do Experts and Naive Observers Judge Printing Quality Differently," Proc. Image Quality and System Performance, SPIE, Vol. 5294, 2004, pp. 132-145.

[10] Tin Kam Ho and Henry S. Baird, "Large-Scale Simulation Studies in Image Pattern Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 10, October 1997, pp. 1067-1079.

[11] Thomas P. Hogan, "Psychological Testing, a Practical Introduction," John Wiley and Son, 2003, p. 232.

[12] Theo Pavlidis, "Effects of Distortions on the Recognition Rate of a Structural OCR System," Proc. IEEE Computer Vision and Pattern Recognition Conference, Washington, DC, 21-23 June 1983, pp. 303-309.

[13] Steve Rice, Frank Jenkins, Thomas Nartker, "The fourth annual test of OCR accuracy," UNLV Information Science Research Institute, December 1994.

0-7695-2822-8/07 $25.00 © 2007 IEEE