Technical Report 56

Automatic Estimation of the Legibility of Binarised Mixed Handwritten and Typed Documents

Martin Stommel and Gideon Frieder
TZI, Universität Bremen

TZI-Bericht Nr. 56
2010

Universität Bremen

TZI-Berichte

Herausgeber:
Technologie-Zentrum Informatik und Informationstechnik
Universität Bremen
Am Fallturm 1
28359 Bremen
Telefon: +49-421-218-7272
Fax: +49-421-218-7820
E-Mail: info@tzi.de
http://www.tzi.de

Automatic Estimation of the Legibility of Binarised Mixed Handwritten and Typed Documents

M. Stommel[1], G. Frieder[2]

[1]Artificial Intelligence Group, TZI, University of Bremen, Germany.
[2]Engineering Management and Systems Engineering,
The George Washington University, Washington, DC 20052.
Email: mstommel@tzi.de, gfrieder@gwu.edu

Abstract

Document enhancement tools are a valuable help in the study of historic documents. Given proper filter settings, many effects that impair the legibility can be evened out (e.g. washed out ink, stained and yellowed paper). However, because of differing authors, languages, handwritings, fonts and paper conditions, no single parameter set fits all documents. Therefore, the parameters are usually tuned in a time-consuming manual process to every individual document. To simplify this procedure, this paper introduces a classifier for the legibility of an enhanced historic text document. Experiments on the binarisation of a set of documents from 1938 to 1946 show that the classifier can be used to automatically derive robust filter settings for a variety of documents.

Keywords: Document analysis, binarisation, handwritten, legibility

1 Introduction

Readers of historic documents are often confronted with damaged or degraded pages where the text has become difficult to read. Stains and washed out ink impair the contrast of the text and hinder the efficient study of bigger numbers of documents.

A certain improvement can be achieved by the application of digital image processing techniques: Shading filters achieve a homogeneous background intensity. Smoothing filters remove noise, sharpening filters accentuate details and contours. By adjusting the filter parameters properly, these methods achieve good results with most types of documents [1]. The downside is that the optimal parameter settings are specific to individual documents. A manual parameter tuning for every document is time-consuming and compensates the advantage of improved legibility.

This paper presents therefore a method to evaluate certain parameter settings of interest automatically. To this end, a classifier is trained that judges the legibility of binarised text documents. The classifier distinguishes between three classes of legibility. Good experimental results are achieved for the diaries of Dr. A. A. Frieder [2].

2 Problem Statement and Related Work

In this paper, the legibility of a text is primarily judged on the basis of binary images. This does not necessarily implicate that binary images are optimal for reading, though contrast enhancement is certainly crucial. Instead, binarised documents are considered as intermediate representation that indicate which pixels belong to the text and which to background. A text is therefore considered as legible if the text pixels form clearly outlined letters without spurious text pixels in background areas. In contrast, a text is considered illegible if letters disintegrate into separate parts and pixels, or if letters disappear completely, or if letters fuse with background areas that are spuriously classified as text. The approach does not prohibit filtering but provides beneficial additional information to filtering methods.

The problem can be stated therefore as finding a parameter setting for a binarisation method that achieves a good legibility for a variety of documents. The notion of optimality in a strict mathematical sense is avoided here, since the legibility can only be estimated with a certain accuracy. The accuracy is limited by the subjective aspects of the topic. However, missing parts of a text can be identified without any doubt.

The difficulty of the binarisation problem depends on the image material. The chosen database [2] is unconstrained in terms of the following visual properties:

- Multiple languages, e.g. Slovak, German
- Multiple authors, e.g. letters from other people
- Handwritten, typed and printed texts, sometimes on the same page
- Sticked in texts, e.g. from newspapers
- Sticked in photos
- Overlay of multiple writings
- Corrections and footnotes
- The back side of a page may shine through
- Multiple fonts on one page
- Washed out and bleeding ink and toner
- Different text colours
- Different fonts, sizes, thickness
- Varying priority for small details
- Text on chequered or lined paper
- Yellowed paper and stains

There are many related topics in document image processing. Optical Character Recognition (OCR) deals with the extraction of features that allow for the segmentation and classification of single letters or even words or phrases. While legibility is a part of the problem, the approach can only be applied to typewritten text. Well-performing OCR-features like e.g. Fourier descriptors represent precise shapes of letters. Because of the mixture of different handwritings and typewritten or printed fonts, such features cannot be generalised easily.

Handwriting Analysis and Signature Recognition are limited with respect to typed text. Dynamic features from Signature Recognition based on the movement of the pen are not available. Varying authors complicate the problem.

Methods from Document Structure Analysis provide a useful segmentation of a page into separate text and image sections. The optimisation of legibility is not the primary goal of the method, though.

3 Automatic Legibility Estimation

The core idea is to produce multiple binary images from a document using different parameter settings and choose the best result based on an automatic legibility estimation. To this end, a legibility function $\Lambda(\phi) : \mathbb{R}^d \mapsto \mathbb{R}$ must be found that judges a parameter set ϕ. However, the parameter set ϕ alone does not contain information on the legibility of a text. To incorporate this information, a set ϕ is characterised by a d-dimensional feature descriptor ψ that is computed on the resulting binary image. The return value of the function is a scalar value that indicates the legibility. The legibility could be measured as the reading speed (of a human) for some sample document or the number of errors during reading. However, a number of discussions and psychological experiments are required to find reliable samples and units. Therefore, in this paper, a simplification is conducted. To check if the approach is viable, the legibility is stated in terms of a natural number based on the author's personal aesthetic sensation:

- The grade 0 is given if the binarisation result does not represent legible text, e.g. black or white areas, photos, blobs and scattered points.
- The grade 1 is given if the binarised image represents text, though letters may be fragmented or merged.
- The grade 2 is given if the binarised text is clearly outlined or if no better binarisation can be achieved.

The legibilty function therefore has the simplified form

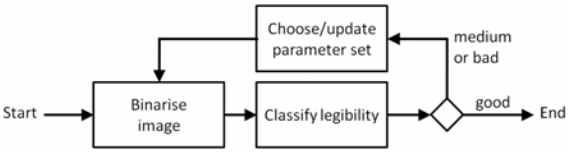
$$\Lambda(\psi) : \mathbb{N}^d \mapsto \{0, 1, 2\}. \tag{1}$$

The mapping from the features ψ to the legibility Λ is learned from sample data.

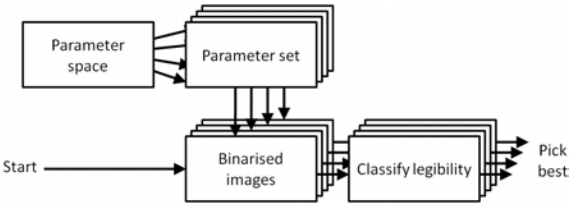
The function Λ can be used to iteratively optimise an initial set of binarisation parameters, e.g. using nested intervals or gradient descent (ascent) (cf. fig. 1a). However, in face of the coarse quantisation of the result value, a parallel approach seems favourable where all parameter sets of interest are checked exhaustively (fig. 1b).

To account for stains and uneven background intensities it seems also advantageous to subdivide a bigger document into smaller sub-regions and optimise the parameter sets locally. The resulting parameter sets can be compared to the results of adjacent regions and filtered if necessary.

The next section discusses the features that represent the domain of the legibility function.



a) Iterative document processing



b) Parallel document processing

Figure 1: Iterative or parallel document processing

4 Domain of the Legibility Function

The feature extraction is conducted on binarised image regions of 100×100 pixels size. The regions size is a trade-off between the number of letters in the region and the spatial frequency of the adaptation to the background gradient. Although the font sizes vary strongly, the regions are big enough to contain whole letters or words.

For a binary image, the number of connected black regions b_r, white regions w_r, the number of black pixels b_p are computed. From these values, the first components of a 14-dimensional feature vector $\psi = (\psi_0, \psi_1, \ldots, \psi_{13})$ are computed as

$$\begin{aligned} \psi_0 &= b_p/\text{region size}, \\ \psi_1 &= b_r/\text{region size}, \\ \psi_2 &= w_r/\text{region size}. \end{aligned} \tag{2}$$

These features represent the proportion of black pixels in the region, as well as a measure of the fragmentation into black and white sub-regions. The aim of these features is to indicate broken lines, isolated points and noise. The next features represent averaged properties of connected black regions, that is to say

$$\begin{aligned} \psi_3 &= \text{mean area of a connected black region}, \\ \psi_4 &= \text{mean width of the surrounding square}, \\ \psi_5 &= \text{most frequent principal orientation}, \\ \psi_6 &= \text{mean ratio of area to width}. \end{aligned} \tag{3}$$

To incorporate information on the stroke thickness of a letter, the binarised region is eroded using a 3×3-box. The features $\psi_7 \ldots \psi_{13}$ are computed the same way as for $\psi_0 \ldots \psi_6$, just on the eroded image.

Fourier descriptors and other shape features used for OCR (see [3] for a review) are powerful for the distinction of different letters. However, in face of different languages, writers, handwritings and fonts, they seem too selective for the present task.

Table 1: Confusion matrix for the test of the legibility function using a 14-dimensional feature descriptor.

Recognised legibility	Annotated legibility		
	good	medium	bad
good	158	31	9
medium	19	60	15
bad	5	60	428
Precision	0.70	0.64	0.92
Recall	0.87	0.40	0.95

Please note that the features do not computationally depend on the size of the region. Complete scale invariance, however, seems unlikely because effects of the general document structure may become more important at higher scales.

A SVM-classifier [4] with soft margin is used to learn the mapping of $\psi \mapsto \{0, 1, 2\}$ from a set of sample images. To this end, a set of 18 documents from the data base [2] is chosen. For each document, 10 regions are randomly selected and binarised using the thresholds 30, 74, 101, 117, 128, 139, 155, 182 and 226. The resulting 1620 binary images are manually annotated, randomised and split into approximately equally sized training and test sets.

A second classifier is trained under the same conditions except that only features ψ_0, ψ_1 and ψ_3 are used.

The structure of the learned mapping is now analysed.

5 Value of the Legibility Function

For the 14-dimensional feature descriptor, an accuracy of 86% for the training samples and 82% for the test samples is achieved. Figure 1 shows the confusion matrix. The best results can be achieved for good and bad binarisations, whereas the samples of medium quality are often confused with the other classes. Taking into account the numerical predominance of bad quality samples, confusions with the good quality samples occur more frequently. This indicates a possible overlap of the good and medium class in feature space. The scatter plot in fig. 2 supports this.

The restriction to the first three features brings very similar results. Here, an accuracy of 84% during training and 81% during test is achieved. The tested shape features seem to contribute only little to the results.

The additional features of the opened binary images contribute indirectly to the results. As fig. 3 shows, good and medium quality samples differ mostly in their variance. The low improvement from three to

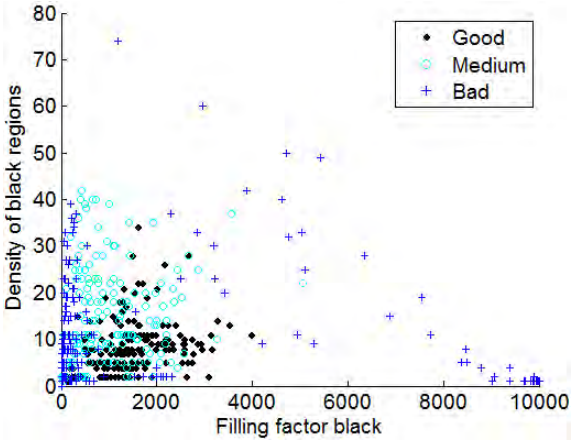


Figure 2: Distribution of ψ_0 (filling factor, black) and ψ_1 (density of black regions) for the test samples including annotation.

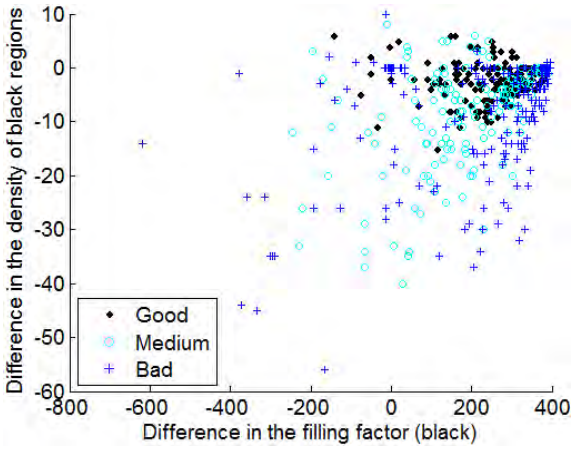


Figure 3: Distribution of the feature differences $\psi_7 - \psi_0$ (filling factor, black) and $\psi_8 - \psi_1$ (density of black regions).

14 features suggests that this difference in variance concerns only samples that have have already been classified correctly using only three features.

Figure 4 shows some of the test samples ordered by classification result. The samples classified as good contain mostly clearly outlined letters but sometimes also letters with small holes, or straight lines from the borders of sticked in papers.

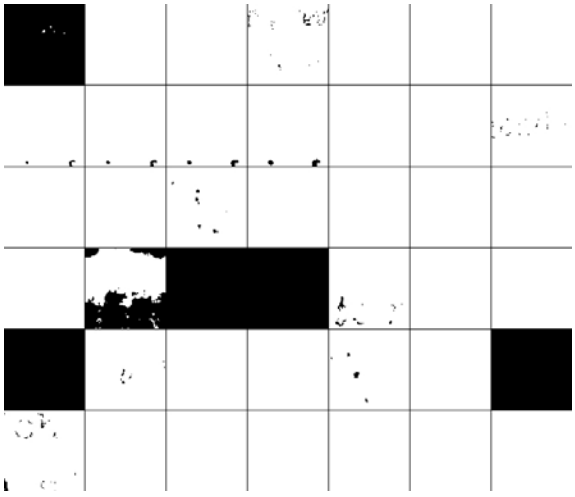
Samples classified as medium contain mostly letters with thin strokes, broken lines and sometimes parts of photos. They are often visually close to the good quality samples.

The image regions classified as bad contain primarily bright samples with occasional clutter, as well as dark parts of photos.

In summary, the classification results seem reasonably good. Since the intercoder reliability has not been measured yet, it is unclear if the samples are suited to evaluate better classifiers. Given the subjective aspects of the topic, different people will

a) Binary images classified good

b) Binary images classified medium



c) Binary images classified bad

Figure 4: Test samples and output of the legibility classifier. For every class, 60 binary images are shown.

probably judge the legibility of a sample different.

The next section shows an example for a parameter optimisation based on the trained legibility function.

6 Parameter Optimisation and Filtering

For one of the optimisation strategies to be viable, the legibility function must have a robustly detectable maximum over the space of the parameters ϕ. To check if this is the case, a simple binarisation system is introduced.

As before, the document is subdivided into overlapping regions of 100×100 pixels. For every region, a set of binary images is produced by thresholding at all values from 0 to 255. Every binary image is then rated by the legibility function.

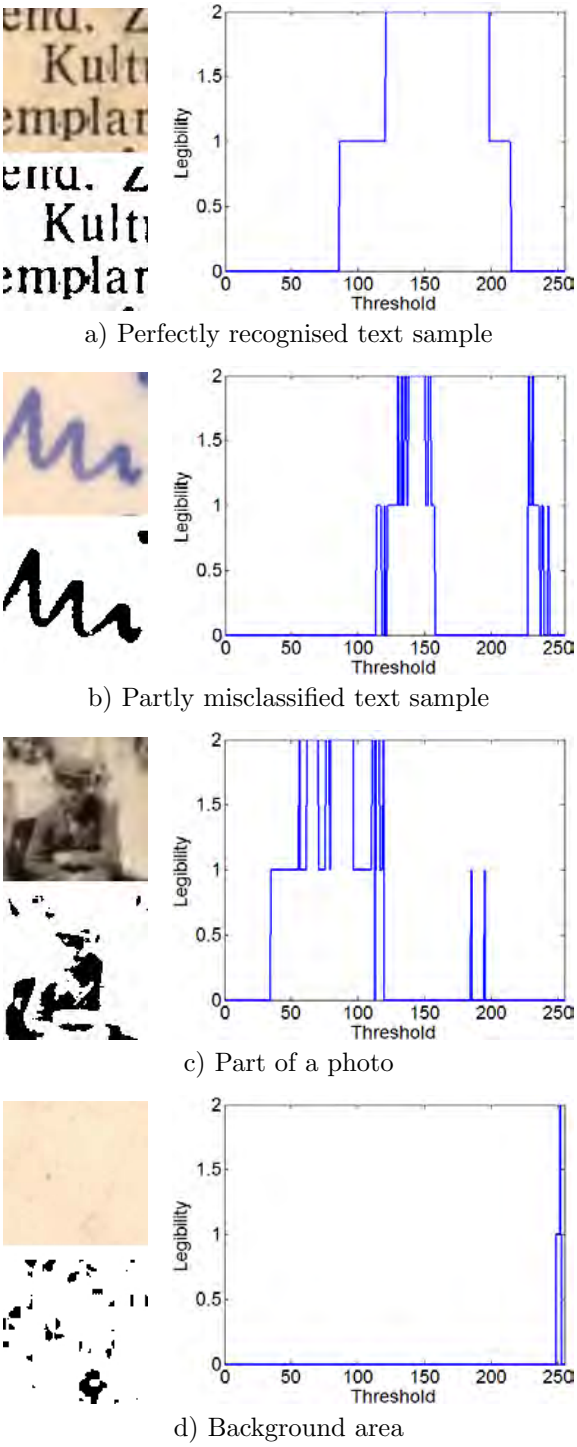
Figure 5 plots the computed legibility versus the threshold for a few example regions. The first plot (a) shows the legibility for text region. The output of the function has a single broad maximum indicating a wide range of viable parameters. The center of gravity of the function is computed and used to threshold the original region. The result is a perfectly legible binarised image section.

The second plot (b) shows a less robust example where the parameter interval 157–226 has not been classified as legible. Nevertheless, the center of gravity provides a good binarisation result. The span from the minimum value classified as legible and the maximum value classified as legible is as big as in the previous example.

Plot (c) shows the output of the legibility function if the input is not a text region but a photo. It turns out that many of the resulting binary images have statistical properties that are similar to binarised text. A wide range of thresholds is marked as suitable for binarisation. The function cannot reliably distinguish between text and photos.

Plot (d) shows a typical result for a background region. The function identifies only a very small range of values as suitable for binarisation. The resulting binary images show usually noise, Jpeg-artefacts and paper edges. The tested shape features do not distinguish between letters and Jpeg-artefacts.

To summarise, the output of the legibility function is usually suited to find a good binarisation threshold. In noisy cases, the center of gravity serves as a robust candidate. While graphical areas cannot be distinguished from text areas, the recognition of empty background regions seems possible. In these



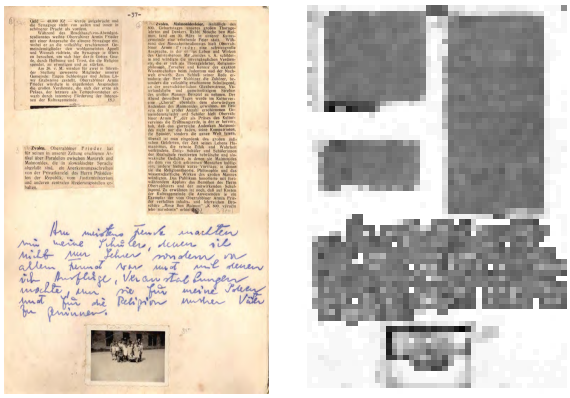
a) Perfectly recognised text sample

b) Partly misclassified text sample

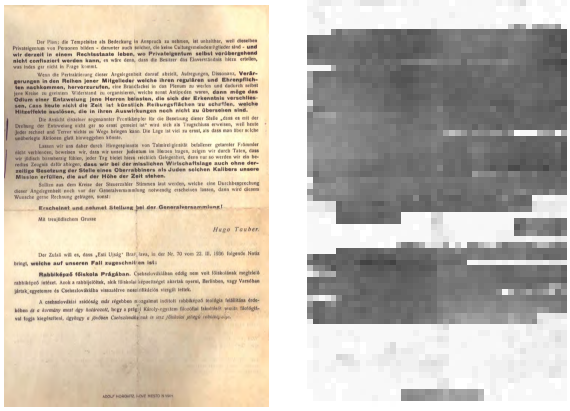
c) Part of a photo

d) Background area

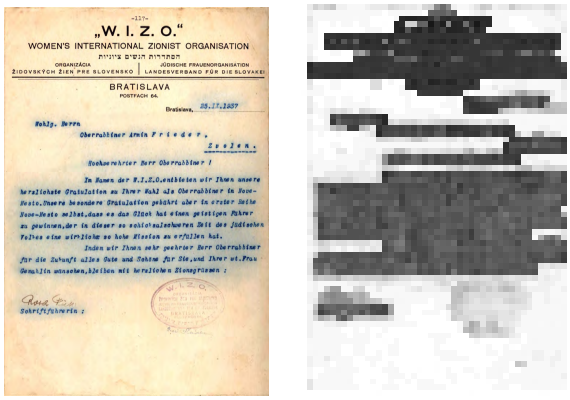
Figure 5: Dependency of the legibility on the binarisation threshold for different image material. The bitmaps on the left show the original gray value image, as well as a binarised version thresholded at the center of mass of the legibility curve on the right.



a) Sample document "M.5_191_38"



b) Sample document "M.5_191_112"



c) Sample document "M.5_191_122"

Figure 6: Text and background estimation based on the range of thresholds classified as good. Dark gray indicates a big range of good binarisation results and therefore a high text probability.

areas, only a very small part of the parameter space is classified as legible (cf. fig. 6).

Since artefacts of the image compression do not improve the legibility of the document, the threshold is set to zero if the range of good parameters is too small. This results in white non-text/non-photo areas.

In order to binarise a full document, the computed thresholds are compared with those of adjacent regions and filtered by a median filter. This reduces potential misclassifications. To keep the borders of the regions invisible, the threshold of every pixel in the document is interpolated bilinearly between the thresholds of the four neighbouring regions. Close to text regions the background detection is suppressed to avoid a fading of the text as a result of the interpolation.

Figure 7, 8 and 9 show the binarisation results for three documents. The legibility of the example in figure 7 is very good for the printed and the handwritten bold text. However, not all of the thin pencil marks are present in the binarised document. Since they are very close to the dark, bold text or the photo, the binarisation threshold is tuned to the dominating neighbouring structures. This can be compensated to a certain degree by a sharpening operation, though this is not the focus of this paper.

The example in fig. 8 seems complete and very clear. The text shining through from the backside of the original document is suppressed.

The example in fig. 9 is good except from the left text border and the stamp on the lower right. The inner part of the stamp is spuriously classified as background and faded out. A sharpening operation between the parameter optimisation and the actual thresholding lets the letters stand out much clearer but also emphasises background noise.

7 Conclusion

This paper presents a classifier for the legibility of binarised text. Since the method does not explicitly perform character recognition, it is applicable to both typewritten and handwritten text. The time-consuming step of manual filter tuning can be omitted, because all parameters are found automatically.

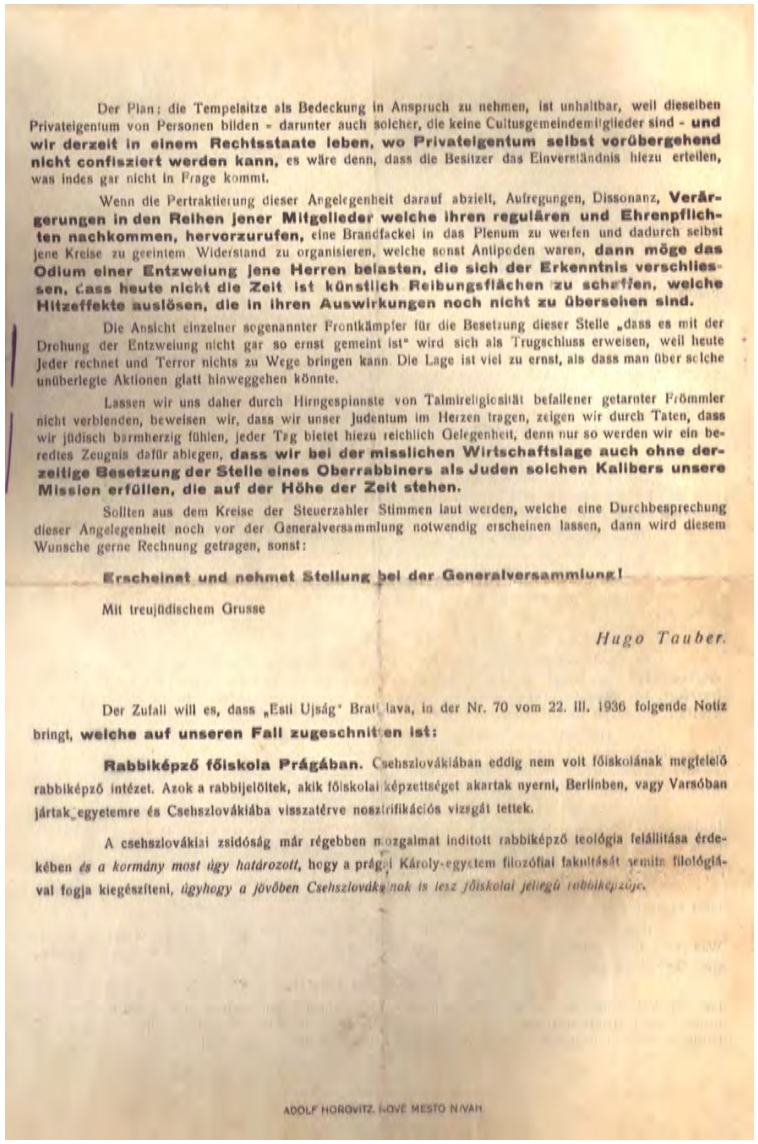
The experiments prove that the method is suited to automatically explore the domain of filter parameters and identify a stable subset. An accuracy of 82% has been achieved for a test set of binarised image regions.

With the present features, however, fine automatic considerations of competing good parameterisations are constrained by a certain overlap in feature space.

8 Acknowledgements

The presented work is a part of the joint research activities by several members of our group. The authors would therefore like to thank Andree Luedtke for sharing his experience in OCR, Arne Jacobs for sharing his experience in the detection of salient

Figure 7: Original and binarised version of sample document "M.5_191_38"



Der Plan: die Tempelsitze als Bedeckung in Anspruch zu nehmen, ist unhaltbar, weil dieselben Privateigentum von Personen bilden - darunter auch solcher, die keine Cultusgemeindemitglieder sind - **und wir derzeit in einem Rechtsstaate leben, wo Privateigentum selbst vorübergehend nicht confisziert werden kann,** es wäre denn, dass die Besitzer das Einverständnis hiezu erteilen, was indes gar nicht in Frage kommt.

Wenn die Pertraktierung dieser Angelegenheit darauf abzielt, Aufregungen, Dissonanz, **Verärgerungen in den Reihen jener Mitglieder welche ihren regulären und Ehrenpflichten nachkommen, hervorzurufen,** eine Brandfackel in das Plenum zu werfen und dadurch selbst jene Kreise zu geeintem Widerstand zu organisieren, welche sonst Antipoden waren, **dann möge das Odium einer Entzweiung jene Herren belasten, die sich der Erkenntnis verschliessen, dass heute nicht die Zeit ist künstlich Reibungsflächen zu schaffen, welche Hitzeffekte auslösen, die in ihren Auswirkungen noch nicht zu übersehen sind.**

Die Ansicht einzelner sogenannter Frontkämpfer für die Besetzung dieser Stelle „dass es mit der Drohung der Entzweiung nicht gar so ernst gemeint ist" wird sich als Trugschluss erweisen, weil heute Jeder rechnet und Terror nichts zu Wege bringen kann. Die Lage ist viel zu ernst, als dass man über solche unüberlegte Aktionen glatt hinweggehen könnte.

Lassen wir uns daher durch Hirngespinnste von Talmireligiosität befallener getarnter Frömmler nicht verblenden, beweisen wir, dass wir unser Judentum im Herzen tragen, zeigen wir durch Taten, dass wir jüdisch barmherzig fühlen, jeder Tag bietet hiezu reichlich Gelegenheit, denn nur so werden wir ein beredtes Zeugnis dafür ablegen, **dass wir bei der misslichen Wirtschaftslage auch ohne derzeitige Besetzung der Stelle eines Oberrabbiners als Juden solchen Kalibers unsere Mission erfüllen, die auf der Höhe der Zeit stehen.**

Sollten aus dem Kreise der Steuerzahler Stimmen laut werden, welche eine Durchbesprechung dieser Angelegenheit noch vor der Generalversammlung notwendig erscheinen lassen, dann wird diesem Wunsche gerne Rechnung getragen, sonst:

Erscheinet und nehmet Stellung bei der Generalversammlung!

Mit treujüdischem Grusse

Hugo Tauber.

Der Zufall will es, dass „Esti Ujság" Brat[illegible]lava, in der Nr. 70 vom 22. III. 1936 folgende Notiz bringt, **welche auf unseren Fall zugeschnitten ist:**

Rabbiképző főiskola Prágában. Csehszlovákiában eddig nem volt főiskolának megfelelő rabbiképző intézet. Azok a rabbijelöltek, akik főiskolai képzettséget akartak nyerni, Berlinben, vagy Varsóban jártak egyetemre és Csehszlovákiába visszatérve noszrifikációs vizsgát tettek.

A csehszlovákiai zsidóság már régebben mozgalmat inditott rabbiképző teológia felállitása érdekében *és a kormány most úgy határozott,* hogy a prágai Károly-egyetem filozófiai fakultását jemitz filológiával fogja kiegészíteni, *úgyhogy a jövőben Csehszlovákiának is lesz főiskolai jellegü rabbiképzője.*

ADOLF HOROVITZ, NOVÉ MESTO N/VAH.

Der Plan: die Tempelsitze als Bedeckung in Anspruch zu nehmen, ist unhaltbar, weil dieselben Privateigentum von Personen bilden - darunter auch solcher, die keine Cultusgemeindemitglieder sind - **und wir derzeit in einem Rechtsstaate leben, wo Privateigentum selbst vorübergehend nicht confisziert werden kann,** es wäre denn, dass die Besitzer das Einverständnis hiezu erteilen, was indes gar nicht in Frage kommt.

Wenn die Pertraktierung dieser Angelegenheit darauf abzielt, Aufregungen, Dissonanz, **Verärgerungen in den Reihen jener Mitglieder welche ihren regulären und Ehrenpflichten nachkommen, hervorzurufen,** eine Brandfackel in das Plenum zu werfen und dadurch selbst jene Kreise zu geeintem Widerstand zu organisieren, welche sonst Antipoden waren, **dann möge das Odium einer Entzweiung jene Herren belasten, die sich der Erkenntnis verschliessen, dass heute nicht die Zeit ist künstlich Reibungsflächen zu schaffen, welche Hitzeffekte auslösen, die in ihren Auswirkungen noch nicht zu übersehen sind.**

Die Ansicht einzelner sogenannter Frontkämpfer für die Besetzung dieser Stelle „dass es mit der Drohung der Entzweiung nicht gar so ernst gemeint ist" wird sich als Trugschluss erweisen, weil heute Jeder rechnet und Terror nichts zu Wege bringen kann. Die Lage ist viel zu ernst, als dass man über solche unüberlegte Aktionen glatt hinweggehen könnte.

Lassen wir uns daher durch Hirngespinnste von Talmireligiosität befallener getarnter Frömmler nicht verblenden, beweisen wir, dass wir unser Judentum im Herzen tragen, zeigen wir durch Taten, dass wir jüdisch barmherzig fühlen, jeder Tag bietet hiezu reichlich Gelegenheit, denn nur so werden wir ein beredtes Zeugnis dafür ablegen, **dass wir bei der misslichen Wirtschaftslage auch ohne derzeitige Besetzung der Stelle eines Oberrabbiners als Juden solchen Kalibers unsere Mission erfüllen, die auf der Höhe der Zeit stehen.**

Sollten aus dem Kreise der Steuerzahler Stimmen laut werden, welche eine Durchbesprechung dieser Angelegenheit noch vor der Generalversammlung notwendig erscheinen lassen, dann wird diesem Wunsche gerne Rechnung getragen, sonst:

Erscheinet und nehmet Stellung bei der Generalversammlung!

Mit treujüdischem Grusse

Hugo Tauber.

Der Zufall will es, dass „Esti Ujság" Brat[illegible]lava, in der Nr. 70 vom 22. III. 1936 folgende Notiz bringt, **welche auf unseren Fall zugeschnitten ist:**

Rabbiképző főiskola Prágában. Csehszlovákiában eddig nem volt főiskolának megfelelő rabbiképző intézet. Azok a rabbijelöltek, akik főiskolai képzettséget akartak nyerni, Berlinben, vagy Varsóban jártak egyetemre és Csehszlovákiába visszatérve noszrifikációs vizsgát tettek.

A csehszlovákiai zsidóság már régebben mozgalmat inditott rabbiképző teológia felállitása érdekében *és a kormány most úgy határozott,* hogy a prágai Károly-egyetem filozófiai fakultását jemitz filológiával fogja kiegészíteni, *úgyhogy a jövőben Csehszlovákiának is lesz főiskolai jellegü rabbiképzője.*

ADOLF HOROVITZ, NOVÉ MESTO N/VAH.

Figure 8: Original and binarised version of sample document "M.5_191_112"

-117-

„W. I. Z. O."

WOMEN'S INTERNATIONAL ZIONIST ORGANISATION

הסתדרות הנשים ציוניות

ORGANIZÁCIA ŽIDOVSKÝCH ŽIEN PRE SLOVENSKO | JÜDISCHE FRAUENORGANISATION LANDESVERBAND FÜR DIE SLOVAKEI

BRATISLAVA
POSTFACH 64.

Bratislava, 25.IX.1937

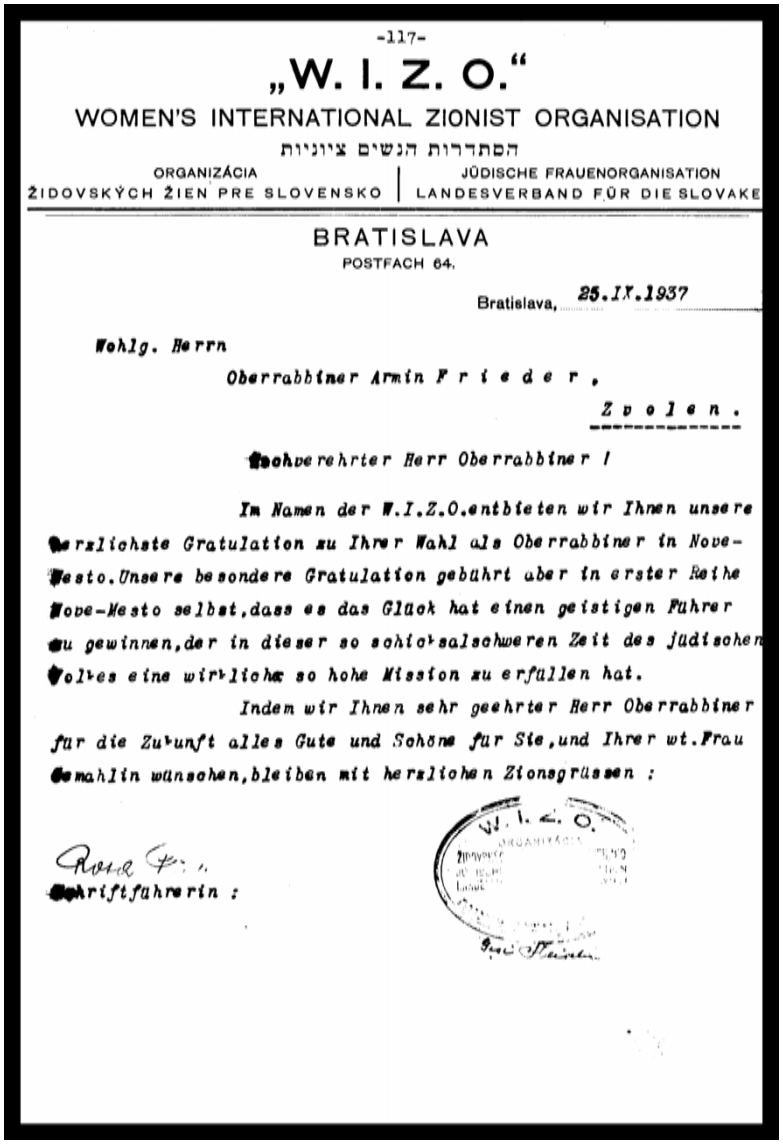
Wohlg. Herrn
Oberrabbiner Armin F r i e d e r ,
Z v o l e n .

Hochverehrter Herr Oberrabbiner !

Im Namen der W.I.Z.O.entbieten wir Ihnen unsere herzlichste Gratulation zu Ihrer Wahl als Oberrabbiner in Nove-Mesto.Unsere besondere Gratulation gebührt aber in erster Reihe Nove-Mesto selbst,dass es das Glück hat einen geistigen Führer zu gewinnen,der in dieser so schicksalschweren Zeit des jüdischen Volkes eine wirklich so hohe Mission zu erfüllen hat.

Indem wir Ihnen sehr geehrter Herr Oberrabbiner für die Zukunft alles Gute und Schöne für Sie,und Ihrer wt.Frau Gemahlin wünschen,bleiben mit herzlichen Zionsgrüssen :

Schriftführerin :



-117-

„W. I. Z. O."

WOMEN'S INTERNATIONAL ZIONIST ORGANISATION

הסתדרות הנשים ציוניות

ORGANIZÁCIA ŽIDOVSKÝCH ŽIEN PRE SLOVENSKO | JÜDISCHE FRAUENORGANISATION LANDESVERBAND FÜR DIE SLOVAKE

BRATISLAVA
POSTFACH 64.

Bratislava, 25.IX.1937

Wohlg. Herrn
Oberrabbiner Armin F r i e d e r ,
Z v o l e n .

Hochverehrter Herr Oberrabbiner !

Im Namen der W.I.Z.O.entbieten wir Ihnen unsere herzlichste Gratulation zu Ihrer Wahl als Oberrabbiner in Nove-Mesto.Unsere besondere Gratulation gebührt aber in erster Reihe Nove-Mesto selbst,dass es das Glück hat einen geistigen Führer zu gewinnen,der in dieser so schicksalschweren Zeit des jüdischen Volkes eine wirklich so hohe Mission zu erfüllen hat.

Indem wir Ihnen sehr geehrter Herr Oberrabbiner für die Zukunft alles Gute und Schöne für Sie,und Ihrer wt.Frau Gemahlin wünschen,bleiben mit herzlichen Zionsgrüssen :

Schriftführerin :

Figure 9: Original and binarised version of sample document "M.5_191_122"

structures, Daniel Moehlmann for his clear assessment, Lothar Meyer-Lerbs, Bjoern Gottfried and Jannis Stoppe for their insights in the processing of non OCR-suited documents, and Otthein Herzog for raising the question.

References

[1] G. Frieder, A. Luedtke, and A. Miene, "Enhancement of document readability," Technologie-Zentrum Informatik und Informationstechnik der Universitt Bremen, Tech. Rep. 43-2007, May 2006.

[2] A. A. Frieder, "The Diaries of Rabbi Dr. Avraham Abba Frieder," 2010, available online at: http://ir.iit.edu/collections/frieder_diaries_README.html. Original is kept as a permanent loan in the Archives of Yad Va Shem, under reference numbers M.5 191,192,193 and 194.

[3] O. D. Trier, A. K. Jain, and T. Taxt, "Feature Extraction Methods For Character Recognition - A Survey," Pattern Recognition, vol. 29, no. 4, pp. 641–662, 1996.

[4] V. Vapnik, The Nature of Statistical Learning Theory. New York: Springer Verlag, 1995.