# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

185,000

200M

Downloads

154
Countries delivered to

Our authors are among the

 $\mathsf{TOP}\:1\%$ 

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us? Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.

For more information visit www.intechopen.com



# Generating Training Sets for the Automatic Recognition of Handwritten Documents

Gabriel Pereira e Silva and Rafael Dueire Lins

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772//52074

#### 1. Introduction

Handwritten character recognition is a task of high complexity even for humans, sometimes. People have different writing "style", which may vary according to psychological state, the kind of document written, and even physical elements such as the texture of the paper and kind of pencil or pen used. Despite such wide range of variation possibilities, some elements tend to remain unchanged in a way that other people, in general, can recognize one's writing and even identify the authorship of a document. Very seldom one is unable to identify one's own writing. Very seldom someone is unable to identify his own writing.

The basis for pattern recognition rests on two corner stones. The first one is to find the minimal set of features that presents all maximum diversity within the universe of study. The second one is to find a suitable training set that also covers all possible data to be classified. Due to the variation of writing styles between people, one should not expect that a general classifier yields good recognition performance in a general context. Thus, one tends to either have general classifiers for very specific restricted vocabularies (such as digits), or to have personalized recognizers for general contexts. The scope of the present work is the latter. In such context it is a burden and very difficult to generate a good training set to allow the classifier to reach a reasonable recognition rate.

This paper proposes a new approach for the automatic generation of the training set for the handwritten recognizer of a given person. The first step for that is to select a set of documents representative of the author's style. In the Internet one may find several public domain sites with font sets. In particular the site Fontspace [21] offers 282 different cursive font sets for download (e.g Brannboll Small, Jenna Sue, Signerica Fat, The Only Exception, Homemade Apple, Santos Dumont, etc). Figure 1 presents an example of some of them. The key idea presented here is "approximating" the author writing by a cursive typographical font, which is skeletonized and a "standard" training set is generated. Such strategy,



detailed as follows, was adopted with success with documents of the Nabuco bequest [12] and of the Thanatos Project [1].

| absocre HOJKRMMOP                     | ABCLETGHLJKLMNÖJ-                           |  |
|---------------------------------------|---|--|
| OR TUVW XYZakadefg                    | QRSTUV WXYZabedefg                          |  |
| hizklmnopgraturwayz                   | híjklmnoparstuvwxyz                         |  |
| 0/23456789                            | 0123456789                                  |  |
| Santos Dumont by Billy Argel - 2007   | Eutemia I by Bolt Cutter Design - 2008      |  |
| ABEDEFGRIJALMO)                       | ABCDEFGHYKLMIQP                             |  |
| DEST WING Zakode fg                   | QRSTUUU X6 YF0 ahodefq                      |  |
| hijk lungsgestne w/x yz.              | hijklmnopgrstuvruxyz                        |  |
| 0125486769                            | najkammopi contentacy                       |  |
| Olho de Boi by Billy Argel - 2011     | Wedding Nightmare by Billy Argel - 2011     |  |
| ABEDEFGHTJK LMN09                     | ABCDEFLYLJK LMNOP                           |  |
| QR SEU VWD Y Zabedefg                 | 2RS TUVWXY Zabcdefg                         |  |
| hijklmnapgrotuvuvoyz                  | hijklmnopgrstuoroxyz                        |  |
| 0123456789                            | 0/23456789                                  |  |
| · ;; '"!?@#\$%&*(/\)                  | ;;"!?@ <b>#\$%</b> &*{(//\)}                |  |
| Bernard by Philing - 1999             | Discipuli Britannica by Peter Wiegel - 2009 |  |
| de Bille Jahr Kalling                 | abcdef gholg KLMNOP                         |  |
| Gette Will practice                   | QRSJUVWXYZabcdefg                           |  |
| hjikmnestarurur                       | hijklmnopgrstuvwzyz                         |  |
| 0/13/56/19<br>;://12#5#/V             | 0123456789                                  |  |
| Gerards Gold by David Kerkhoff - 2010 | Kathleenie by Robotic Attack Fonts - 1997   |  |

**Figure 1.** Examples of cursive font sets extracted from Fontspace [1]

## 2. The proposed method

The choice of a representative training set together with a good set of features is fundamental for the success of automatic pattern recognition. These two factors are tightly linked to each other and in such a way as to grant good recognition results. To obtain a good training set for handwriting recognition of a given author during a period of time in which the writing features are stable (they changes with age, psychological and health factors, etc.) one has to group together the documents that have similar properties. A subset of them that is representative of the set of documents (in general the size of the training set is about 10% of the size of the whole data "universe") is chosen in such a way as to cover the whole diversity of the documents to be transcribed.

The development of the proposed method starts by using a set of cursive fonts. In the Internet one may find several public domain sites with font sets. In particular the site Fontspace [21] offers 282 different cursive font sets for download.

The central difficulty in generating the training set for handwritten documents is to have a "font set" of a specific author to extract the convenient features for patter matching.

The strategy adopted here is:

- Select a number of documents that is representative of the author.
- Process the documents to:
  - remove digitalization borders that may frame the image,
  - b. correct skew,
  - filter-out back-to-front interference (bleeding),
  - binarize the document.
- Transcribe the document into text by a human reader. 3.
- The user should select a number of a cursive font set that bears "some resemblance" to the original author handwriting.
- The text version of the document is typeset in each of the cursive font sets chosen in 5. step 4 (above).
- All the typeset versions of the document are converted into image. 6.
- The image of the original document is skeletonised and then dilated.
- Segment the image in boxes around each letter (font cases) of the skeletonised and dilated version of the original image and the synthetically generated images.
- Apply a deformation transform to make each font case in the synthetic images coincide with the font case of the skeletonised-dilated version of the original document.
- 10. Extract the features from each document and place in a vector.
- 11. Take the Hamming distance between the feature vectors of the synthetic and original images.
- 12. The font set used to generate the synthetic document which provides the smallest Hamming distance is the one to be used as the training set.

The structural features used for pattern recognition, mentioned in step 10 above, are:

Geometric Moments [15] [9];

- Concavity Measurements [16];
- Shape Representation of Profile [14];
- Distance between barycentre points between two consecutive characters;
- Maximum and minimum heights of two consecutive characters.
- Maximum and minimum distance between concavities of two consecutive characters.

The image filtering operations listed in step 2 were performed using HistDoc v.2.0 environment [13] which offers a wide number of tools for historic document image processing including the several algorithms for the removal of back-to-front interference and binarization. The skeletonization and dilation processes in this work were performed using the filters available in ImageJ [20].

Step 9, performing image vectorization, is important to increase the likelihood between the synthetic and the original documents. Such operation is applied to each character in each synthetically generated image by deforming the bounding-box and the strokes until there is a perfect match between the synthetic and the original one. In this "deformation" process some statistical analysis is performed to infer data about inter character and inter word spacing, line and character skew, inter line separation, etc.

The feature vector of a document brings an account of the basic features of the author calligraphy. The Hamming distance between the feature vectors of the synthetic and original images, which is part of step 11, brings an account of their similarity, and is calculated using the formula:

$$H_w = \sum_{n=1}^{N \text{ features}} \left| f_{on} - f_{sn} \right|$$

where  $f_{on}$  and  $f_{sn}$  are the components of the feature vectors of the original and synthetic images, respectively. The choice of a vector of features such that one could extract "information" about the calligraphic pattern of the author shares some ideas with the work in reference [5]. The font set that provides the smallest Hamming distance to the original set is chosen to synthetically generate the whole training dictionary to the classifier.

In what follows the steps described above are detailed in two files of historical documents: the handwritten letters of Joaquim Nabuco that are about one century old and the hand filled information on the books of pre-printed forms of civil certificates from the state of Pernambuco-Brazil, from mid 20<sup>th</sup> century.

#### 3. Results

The strategy presented above for developing the training set was tested in two sets of documents: letters from Nabuco bequest and death certificates from the Thanatos project [1].

# 3.1. Transcribing Nabuco's letters

The Nabuco Project [12] was an initiative of the second author of this paper. It started in 1991 with the aim to preserve the file of letters of Joaquim Nabuco, a Brazilian statesman,

writer, and diplomat, one of the key figures in the campaign for freeing black slaves in Brazil (b.1861-d.1910). The Nabuco file encompasses over 6,500 documents and about 30,000 pages of active and passive correspondence (including postcards, typed and handwritten letters), a bequest of historical documents of paramount importance to understand the formation of the political and social structure of the countries in the Americas and their relationship with other countries. The letters of Nabuco were catalogued and some of them summarized [2] [4], but the bequest was never fully transcribed. The Nabuco Project is acknowledged as being the pioneering initiative in Latin America to attempt to generate a digital library of historic documents. Figure 2 presents an example of letter in Nabuco bequest, written in a blank sheet of paper without lines, which presents a textured background due to paper aging, a horizontal folding mark in its central part, a light backto-front interference (bleeding) as the letter was written on both sides of the sheet of paper. The image was acquired with an operator driven flatbed scanner, using 200 dpi resolution, in true color. There is no marginal noise (borders) framing the image and its skew is negligible.

To automatically generate the training set for recognizing the handwritten letters from Nabuco file a visual inspection was made to find letters that could represent the whole universe of letters. From the Nabuco file 50 letters were chosen and transcribed by historians, yielding 50 text files, totaling 3,584 words. Twenty-five letters (1,469 words) were used to develop the feature set used for training the classifier, and the remaining ones for ground-truth testing. All the selected documents were processed performing the steps listed in step 2 above, which encompasses marginal border removal, image de-skew, removal of back-to-front interference and binarization. An example of resulting document after filtering and binarization using the HistDoc v.2.0 environment [13] may be found in Figure 3. The image in Figure 2 is skeletonized and then dilated using the filters in ImageJ [20]; the resulting image is presented in Figure 4.

The synthetic image generation is performed by choosing a subset of the cursive fontsets available that resemble the writing of the original document. In the case of Nabuco, the subset selected encompassed 15 of the 282 cursive font sets available in Fontspace [21] (e.g. Brannboll Small, Jenna Sue, Signerica Fat, The Only Exception, Homemade Apple, Santos Dumont, etc) that were closer to the author's writing style during the period of interest. The text of the original document was (human) transcribed into a text file, which was typeset using the choosen cursive fontsets. The text of document in Figure 2 typeset using the fonts in "Signerica Fat" type font is shown in Figure 5. The image shown in Figure 5 is now vectorized and "approximated" to the original skeletonized and dilated image by "deforming" each "letter case" and strokes until matching, as much as possible. The resulting image is presented in Figure 6.

The feature vector of each of the synthetic images "deformed" in such a way to the character case to match the original font case was extracted and the Hamming distance of each of them to the skelotonized-dilated original image was calculated. The image that exhibited the minimum distance was the "Signerica Fat" font set presented in Figure 6.

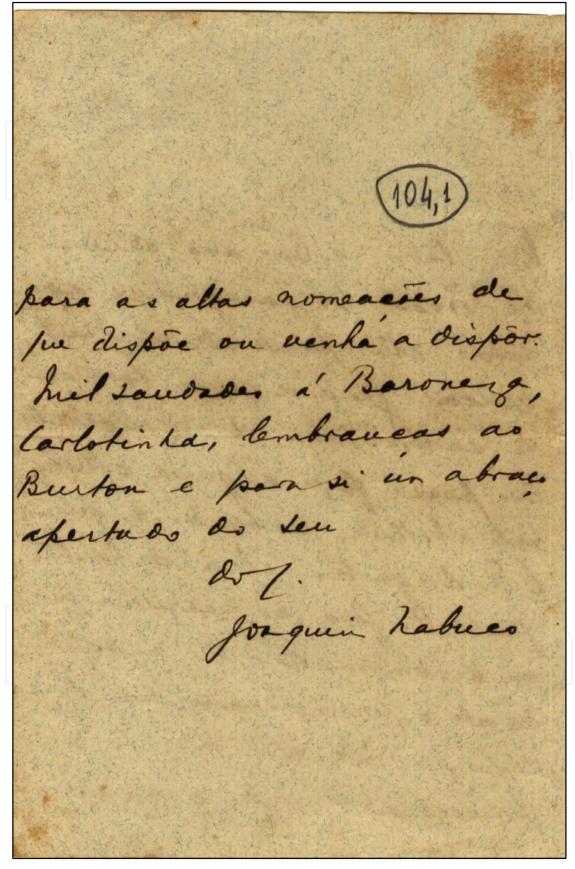


Figure 2. Letter from Nabuco bequest.

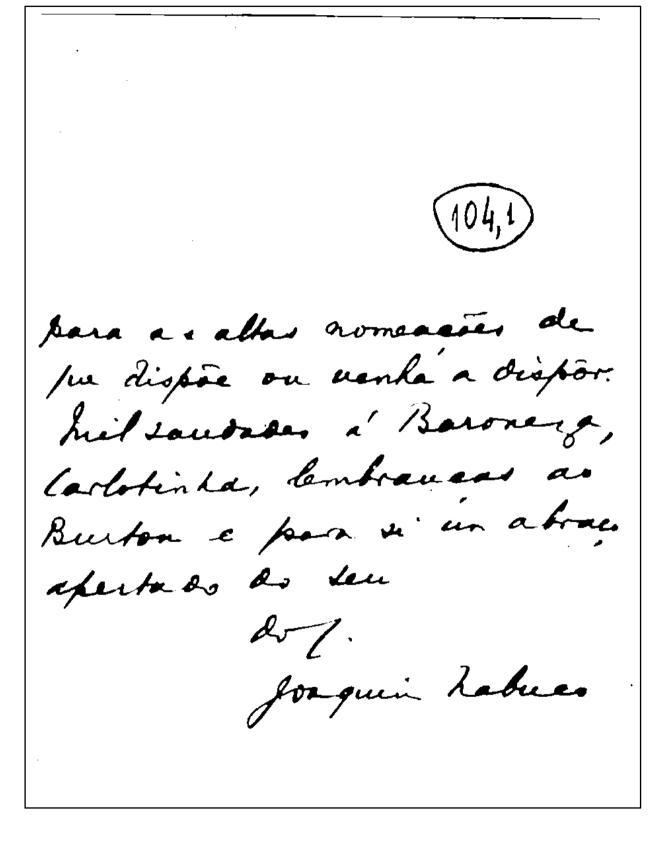


Figure 3. Document shown if Figure 2 after filtering and binarization.

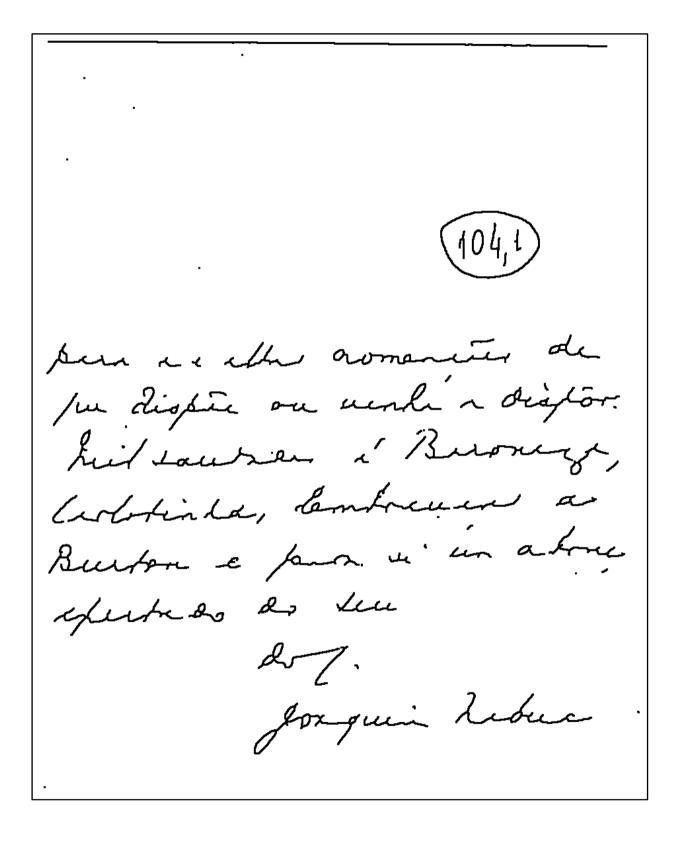


Figure 4. Skeletonized and dilated version of letter shown if Figure 3

104

Para ay altay nomenções de que dispõe ou venha a dispor. Hil jourgoe à Baroneza, Carlotinha, Lembrançay ac Burton a para je um abraça Spertada ja jen Joaquin Rabuco

Figure 5. Synthetic skeletonized image generated from typesetting the text of the original document with the "Signerica Fat" font set.

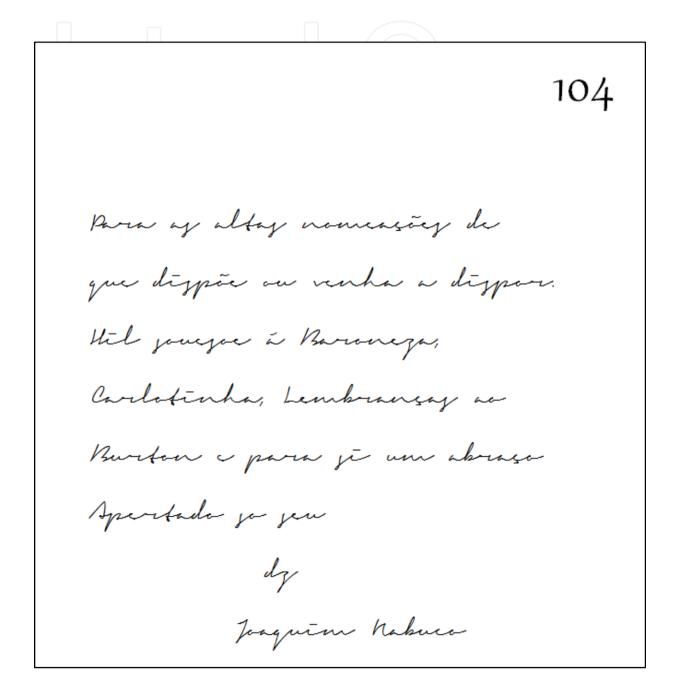


Figure 6. Image of Figure 5 after vectorization and deformation to make a font case matching to the original document after skeletonization and dilatation (Figure 4).

The comparison of the images of Figure 2 and Figure 5 shows several small differences, but there is a mapping path between each letter in the original text (ASCII character) and a "font" that resembles the author calligraphic pattern, which allows the automatic generation of a dictionary of patterns to be used as a training set for the recognizer.

The comparison of the images of Figure 2 and Figure 5 shows several small differences, but there is a mapping path between each letter in the original text (ASCII character) and a "font" that resembles the author calligraphic pattern, which allows the automatic generation of a dictionary of patterns to be used as a training set for the recognizer.

A MLP [8] and two SOM [10] fuzzy classifiers were used in parallel and the majority vote is taken for the transcription of the 25 letters in the document test set, totaling 2,115 words (with at least three letters), both trained with the same dictionary of synthesized words. The result obtained was of 61% (1,294 words) correctly transcribed and 17% (364 words) mismatched into (incorrect) valid words. Testing the whole set of fifty letters (3,584 words), that include the 25 letters used to develop the training set the results were of 67% words correctly transcribed and 15% of "false-positive" words . The result of the classifier applied to the remaining 25 letters yielded a precision and recall of 72%. Table 1 shows the significance of the recognition rate reached may be seen if one attempts to automatically recognize the document in Figure 2 with the classifier trained using the approach presented here and three of the best OCR softwares available today in the market: the Abby FineReader version 12 [19], Omnipage [23] and OCRopus 0.3.1 (alpha3) [22] that calls Tesseract.

Table 1 witnesses the suitability of the melhod proposed here. It is interesting to notice that even the human reader does not know what Joaquim Nabuco meant with the symbol (?) just before his signature. The transcription automatically made using the methodology proposed here may be considered very successful, overall if compared with the transcriptions obtained by the commercial OCRs tested (Tesseract produced no output at all). One interesting fact to observe is that although the grammatically correct accent in the third line of the text is "à", Nabuco's writing was very calligraphically "imprecise" and looks as "á", as automatically transcribed. One may not consider that an error or even that he mispelled the lexeme "à", because the "á" in isolation does not exist in Portuguese. The addition of a dictionary may solve such a problem as well as some other as for instance the transcribed word "Hil" does not exist in Portuguese and the only possible valid candidate is the correct word "Mil" (one thousand).

#### 3.2. Word recognition in death certificates

Death certificates provide important data such as causa mortis, age of death, birth and death places, parental information, etc. Such information may be used to analyze not only what caused the death of the person, but also a large number of demographic information such as internal migration, the relation of death cause with marital status, sex, profession, etc.

Images were acquired by The Family Search International Institute using a camera-based platform.

| 104  | П   | 104   |
|--|---|---|
| Para as altas nomeações de<br>que dispõe ou venha a<br>dispor.<br>Mil saudades à Baroneza,<br>Carlotinha, Lembranças ao<br>Burton e para si um abraço<br>Apertado do seu | en()  | Para as altas nomeações de<br>que dispõe ou venha a<br>dispor.<br>Hil souesoe á Baroneza,<br>Carlotinha, Lembranças ao<br>Burton e para si um abraço<br>Apertado do seu<br>dz |
| Joaquim Nabuco   |   | Joaquim Nabuco  |
| Human transcribed text   |   | Proposed Method   |
| 1*^  | -* * ' ^ ^^"f* ^CjL-îU As<br>À-«-\$tjt_Jt-C |   |
| Omnipage   | Abby FineReader Professional Edition 12     | OCRopus/Tesseract   |

**Table 1.** Human transcribed text of the document in Figure 2 and the automatic transcriptions by the Proposed Method, Omnipage, Abby FineReader and Tesseract.

Thanatos [1] is a platform designed to extract information from the Death Certificate Records in Pernambuco (Brazil), a collection of "books" kept by the local authorities from the 16th century onwards. The current phase of the Thanatos project focuses on the books from the 19th century. During such period, registration books were pre-printed with blank spaces to be filled in by the notary, as shown in Figure 7. Pre-processing is performed to remove noisy borders using the algorithm described in reference [6] incorporated in the

HistDoc Plarform as this step influences all the result of the other subsequent algorithms, the result of which is shown in Figure 8. Image processing continues on the border-removed image (Figure 8) to make image-size (resolution) uniform, binarize, correct skew (using the algorithm Ávila and Lins [3], 2005 also implemented in HistDoc [13]), remove salt-andpepper and clutter noises, and finally splitting an image in two images each of them corresponding to one death certificate as shown in Figure 9.

Notaries in Brazil are a concession of the State. They are a permanent position many people exercise throughout their lives. Thus, most record books are written by a single person, allowing one to use the strategy proposed here to train the classifier to recognize the content of the different fields. Masks are then applied to extract the content of each of the fields filled in by notaries to extract the content.

#### They are:

- $N^{\circ}$  (Register number) placed at the top of the left margin of the register. It conveys numerical information only. Example: No 19.945.
- Data (Date) the date is written in words and the information is filled in three fields for day, month, and year in this sequence. Example: Aos vinte e três dias do mês de janeiro de mil novecentos e sessenta e seis (At the twenty three days of the month of January of one thousand nine hundred and sixty six).
- Nome do cartório (Notary name) this field holds the name of the place where the notary office was found. Example: neste cartório da Encruzilhada (at this notary office at Encruzilhada).
- Município do Cartório (City of the notary office) Example: município de Recife (at the city of Recife).
- Estado do Cartório (State of the notary office) Example: Estado de Pernambuco (State of Pernambuco).
- Nome do Declarante (Name of declarer) Name of who attended the office to inform the death. Example: compareceu Guilherme dos Santos (attended Guilherme dos Santos).
- Nome do Médico (Name of the Medical Doctor) Name of the M.D. who checked the death. Example: exibindo um atestado de óbito firmado pelo doutor José Ricardo (showing a death declaration signed by doctor José Ricardo).
- Causa mortis Specifies the reason of the death in the declaration from the M.D. Example: dando como causa da morte edema pulmonar, o qual fica arquivado (that states as cause of death lung edema, which is filed).

The first strategy reported in reference [1] for information recognition in the Thanatos platform was to transcribe the fields using the commercial OCR tool ABBYY FineReader 12 Professional Editor [19]. The results obtained were zero correct recognition for all fields, including even the numerical ones. Such disappointing results forced the development of a recognition tool for the Thanatos platform based in the approach in reference [17] that makes use of a set of geometrical and perceptual features extracted from "zoning" the image.

"Zoning" may be seen as splitting a complex pattern in several simpler ones [18] [11] [7]. The original Thanatos strategy used dictionaries to analyze the possible "answers" to the blank fields. The original results of tests performed with 300 death certificates extracted from the same book of death records [1] were already considered reasonable and are shown in the first column of Table 2.

The adoption of the strategy presented here to generate the features of the writer through the modification of a cursive type font text was adopted. The list of all cities and places (villages, neighborhoods, etc) in the state of Pernambuco was collected from IBGE (the Brazilian Geographic and Statistical Institute) a social science research institute responsible for demographic and economic statistics and data collection in Brazil. Another list of family names was also generated having as basis the local phone directory. Those lists were "typeset" using the synthetic set of features extracted and then used to train the classifiers. The results obtained adopting this strategy is presented in the New column in Table 2. It is important to stress that the same parallel architecture (MLP + 2 SOM) fuzzy classifiers with majority vote was used in both cases, only with different training sets.

| Field                              | Thanatos | New    |
|------------------------------------|----------|--------|
| Name of Notary                     | 98.0%    | 98.5%  |
| City of the Notary                 | 71.0%    | 94.0%  |
| State of the Notary                | 98.0%    | 100.0% |
| Place of death                     | 31.0%    | 73.0%  |
| Numbers in writing - Time of obit  | 69.0%    | 91.0%  |
| Numbers in writing - Date of death | 69.0%    | 91.0%  |
| Numbers in writing - Date of birth | 69.0%    | 91.0%  |
| Color of skin                      | 100.0%   | 100.0% |
| Marital status                     | 100.0%   | 100.0% |

**Table 2.** Recognition rate for non-numerical fields in 300 certificates.

| 19.945  | · dos linte eties - dias do mês de faccion de mil novecentos  |   |
|---|---|---|
| 1   | Municipio de Perite , neste Carlório da Recuyillande - Municipio de Perite . Endo de Crustimus - compareceu Quillioneus kon parto Celvo Grandt exibindo um alestavo de óbito tirmado polo de lo Carlo |   |
|   | Comparere Quaille on 141 Par  |   |
|   | exibindo um atestado de óbito tiemodo sels doutor Celvo Carando.  |   |
|   | dando como causa da   |   |
|   | morte, Prematinidade de   |   |
|   | o qual fica arquivado, declarou que pro Juntum dade da  | **************************              |
|   | Tuerusiliado.   |   |
|   | às oly - horas e dez - do dia degeno ve - de  |   |
|   | Taleces of files  |   |
|   | do sexo (wasulino, de cor parda comqualis dias  |   |
|   | de vido, natural de Cerriaisteuro-  |   |
|   | Profissão Estado Civil -  |   |
|   | residente   |   |
|   | Tilho de buis leastano de plus e de morta   |   |
|   | Quara da filos, lavadura, naturais  | *************************************** |
|   | ally Estado.  |   |
|   |   |   |
|   | Observações: Les feit às exercios dotering  |   |
|   | hereties ereas.   |   |
|   |   | *************************************** |
|   |   |   |
|   | Para constar, lavrei este termo, depois de ter cumprido todas as formali-   | *************************************** |
|   | dades de acordo a lei vigente, e depois de lido e achado conforme, é assinado. Eu, escrevente jura-   |   |
|   | mentado, escrevi, e eu,   |   |
|   | Oficial do Registro Civil, subscrevo.   | -                                       |
|   | Buffmu das Cant   |   |
|   |   |   |
|   |   |   |
| N.º 19. 946   | Aos Sente ties - dia do mês de formo de mil novecentos  |   |
| 1   | e persente, peis , neste Cartório da que unillocado   |   |
|   | Municipio de Reine . Estado de Servicio luno compareceu Jui Burerere dos partos   |   |
|   | exibindo um atestado de óbito firmado seto de esto, belso brand -   |   |
|   | dando como causa da   |   |
|   | morte, fusturidade.   |   |
|   | o qual fica arguivada, declarou que no funterus dade da   |   |
|   |   |   |
|   | do Vienti quatrioras do dia Viente de   |   |
|   | Tageen Jose da filo .   |   |
|   | Transcent Jan Comment   | *************************************** |
|   | do sexo properliero, de corpardo - com dais dias  |   |
|   | ale vida - natural de seuseusteurs-   |   |
|   | Profissão Estado Civil  |   |
|   |   | *************************************** |
|   | 11/10 de · Otquiel feverira de filos e de felena  | *************************************** |
|   | Illustio de hive desertico realtirais   |   |
|   | Sendo sepullado no cemitério de Varres  |   |
|   |   |   |
|   | Observações: fei feil à 1xdemas de teures   | •                                       |
|   |   |   |
|   |   |   |
|   | Para constar, lavrei este termo, depois de ter cumprido todas as formali-   | *************                           |
|   | dades de afordo a lei vigente, e depois de lido e achado conforme, é assinado.  |   |
|   | Eu, escrevente jura-<br>mentado, escrevi, e eu,   |   |
|   | Olietal do Registro Civil, subscrevo.   |   |
|   | Fullman da Laut   |   |
|   |   |   |
|   |   |   |
| CONTRACTOR OF THE PARTY OF THE |   |   |

Figure 7. Original image from a book of printed forms of death certificates in Pernambuco (Brazil) – 1966.

| N. 19.945.  |  |   |
|-------------|--|---|
| -1          | e persente, seu , neste Carlório da hungelliade -  |   |
|             | Municipio de Rente Estado de terresculuio -  |   |
|             | compareceu quilliarius tor jactor  |   |
|             | exibindo um atestado de óbito sirmado pela dento belvo brandt  |   |
|             | morte, Prematini dade de   |   |
|             |  |   |
|             | o qual fica arguivado, declarou que po husturis dade da  | - |
|             | Tuerusuliado.  |   |
|             | as des - horas e des - do dia descrive - do  |   |
|             | de mil novecentos e pessente , per   |   |
|             | faleced for filos  |   |
|             | do com Quantilian - de câs any da - com Qualis dia   |   |
|             | de sexo hosselino, de côr randa - comangla dias  |   |
|             | Profissão Estado Civil   | - |
|             | residente  |   |
|             |  |   |
|             | silho de buiz leastano, de filos e de marta  |   |
|             | Quario da filos, lavadura, naturais  |   |
|             | Sendo sepullado no cemilério de Varges   |   |
|             | Observações: Le feit à experior doterico   |   |
|             | Special do Estado  |   |
|             |  |   |
|             |  |   |
|             |  |   |
|             | Para constar, lavrei este termo, depois de ter cumprido todas as formali-  |   |
|             | dades de acordo a lei vigente, e depois de lido e achado conforme, & assinado.   |   |
|             | Eu, f = 621 escrevente jura-   |   |
|             | mentalof, escrevi, e eu, Oficial do Registro Civil, subscrevo.   |   |
|             | Just home dan Cant   |   |
|             | Jan  |   |
|             |  |   |
|             |  |   |
| N.º19. 946. | Aos Sente tres - dia do mês de formo de mil novecentos   |   |
|             | at mit noverences  |   |
| 1           | e pessenta, sies -, neste Cartório da Lucuyellacolo  |   |
|             | Municipio de Renge , neste Cartório das Evernyellocado   |   |
|             | municipio de Renge , neste Cartório das Evernyellocolo Municipio de Renge , Estado de Evernyellocolo compareceu Des Brierres dos partos  |   |
| 1           | e personte, vier , neste Cartório das querryellocolo Municipio de Renge , Estado de Erresculturo compareceu feni buerne do fanto pelo brand texibindo um destado de óbito firmado pelo deesto, belso brand t   |   |
|             | municipio de Cerce , neste Cartório das lucrusellocado Municipio de Cerce , Estado de Cercesculturo compareceu Just Prerue dos facetos beles brand t- exibindo um etestado de óbilo firmado celo decitos beles brand t- dando como causa da  |   |
|             | e persente, sees , neste Cartório das prenuellocolo Municipio de Renge , Estado de Cercusulment compareceu General dos facilitos frances exibindo um estestado de óbito firmado cela dento fielas frances transceres dando como causa da morte, presturidade.  |   |
|             | e persente, sees , neste Cartório da premiellocolo Municipio de Renge , Estado de Carron de compareceu fundimente dos factos fellos francit - dando como causa da morte, fundimente de como causa da como como como como como como como com   |   |
|             | e persente, sees , neste Cartório das prenugellocolo Municipio de Renge , Estado de Carriscolo compareceu frui buerne dos fanto como causa da morte, fruaturidade .  o qual fica arguivada, declarou que pro funternidade da Renguirado  |   |
|             | e personte, sein , neste Cartório da prerrigellocolo Municipio de Renge , Estado de Personal de compareceu fundimente dos facilitàs bando como causa da morte, frantizidade .  o qual fica arquivada, declarou que pro proteccidade da Prengillocado do Vierte securioras — do dia Vierte de   |   |
|             | municipio de Renge , neste Gartório da prennellado Municipio de Renge , Estado de Errosanhum comparecen Grund Direccue dos partos bellos brand to dando como causa da morte, praturidade de declaron que por traternadade da lucinidade da lucinidade de distribución de de mil novecentos exercises para de de mil novecentos exercises para la de  |   |
|             | e personte, sein , neste Cartório da prerrigellocolo Municipio de Renge , Estado de Personal de compareceu fundimente dos facilitàs bando como causa da morte, frantizidade .  o qual fica arquivada, declarou que pro proteccidade da Prengillocado do Vierte securioras — do dia Vierte de   |   |
|             | e persente, per , neste Cartório da premiellocolo Municipio de Renge , Estado de Carriscultura compareceu Grusta de Situado de Solito firmado cela dento Bello Grand Texibindo um estestado de óbito firmado cela dento Bello Grand Texibindo como causa da morte, presturidade de declarou que pro funtamendade da Successibilidade de Microscopilidade de mil novecentos exercista pero de Jaseceu forte de filipo -   |   |
|             | e persente, per , neste Cartório da premiellocolo Municipio de Renge , Estado de Carriscultura compareceu Grusta de Situado de Solito firmado cela dento Bello Grand Texibindo um estestado de óbito firmado cela dento Bello Grand Texibindo como causa da morte, presturidade de declarou que pro funtamendade da Successibilidade de Microscopilidade de mil novecentos exercista pero de Jaseceu forte de filipo -   |   |
|             | e persente, peus , neste Cartório da premiellocolo Municipio de Renge , Estado de Persentalment compareceu fundimente dos factos filmando como causa da morte, presturidade de obito firmado cela dento fielas francio dando como causa da morte, presturidade de declarou que pre fundamente de de Victorio de mil novecentos e com deix de la sexo presente perso de de sexo presente de la compara de la sexo presente de la sexo pre |   |
|             | e persenta, seis , neste Gartório das premiellosolo Municipio de Renge , Estado de Personalmento compareceu Guillanerus dos facilitàs de la como causa da morte, presturidante . dando como causa da lucirio fica arguivada, declarou que po protecuidade da lucirio fica arguivada, declarou que po protecuidade da lucirio fica de mil novecentos esperiente, personal de sexo procesa de filia de con causa da matural de consecutivas de la sexo procesa de con dans dias de de con de con de con de con de con de con dans dias de con de c |   |
|             | e persenta, per , neste Gartório da premiellado Municipio de Renge , Estado de Perrescultura compareceu Grusta de Sinta fanta per de   |   |
|             | e persenta, per , neste Gartório da premiellocolo Municipio de Renge , Estado de Personalment compareceu Juni Oriente dos partos personalment exibindo um estestado de óbito firmado esta desito belso brand tomorte, presturidade da dando como causa da morte, presturidade da Cuencipillocado de mil novecentos esperanta pero la faseceu pero de mil novecentos esperanta pero la faseceu pero de filoso e com dans dias profusão , natural de mensentamo Profusão Estado Civil residente  |   |
|             | e persenta, seis , neste Gartório das premiellosolo Municipio de Renge , Estado de Persentalment compareceu Grustando de Silva franto parto de dando como causa da morte, presturidande da dando como causa da l'entregles de de mil novecentos espersenta perso de mil novecentos espersenta perso de materialmento de sexo properso de compareceu de sexo properso de sexo properso de compareceu de filo de compareceu de filos de filos de de filos de fi |   |
|             | e persenta, per , neste Gartório da premiellocolo Municipio de Renge , Estado de Personalment compareceu Juni Oriente dos partos personalment exibindo um estestado de óbito firmado esta desito belso brand tomorte, presturidade da dando como causa da morte, presturidade da Cuencipillocado de mil novecentos esperanta pero la faseceu pero de mil novecentos esperanta pero la faseceu pero de filoso e com dans dias profusão , natural de mensentamo Profusão Estado Civil residente  |   |
|             | e persenta, per , neste Gartório da premiellado de Municipio de Renge , Estado de Persentalment compareceu fundamente dos facilitàs bellos franctitàs de destro de destro de de mil novecentos exercita persenta de de sexo france de filmo de con parte de filmo de con parte de la filmo de con |   |
|             | do sexo magnetino, de côn rando - com dais dia sexo magnete de mil novecentos especiales de sexo de mando como causa da de la fina de de de la descripción de morte, producto de   |   |
|             | e persenta, per , neste Gartório da premisellado de Municipio de Renge , Estado de Persentalment compareceu fundamente dos facilitàs franctitàs de destro bella franctità dando como causa da morte, presturidade de distribución dando como causa da funcionalmente de funcionalmente de mil novecentos exercita persenta de filmo de con para de filmo de con para de filmo de con profissão Estado Civil residente  Sendo sepultado no cemitério de Varges do Ferraço do Servações: La feita a pressura do Servações do S |   |
|             | e persenta, per , neste Gartório da premisellado de Municipio de Renge , Estado de Persentalment compareceu fundamente dos facilitàs franctitàs de destro bella franctità dando como causa da morte, presturidade de distribución dando como causa da funcionalmente de funcionalmente de mil novecentos exercita persenta de filmo de con para de filmo de con para de filmo de con profissão Estado Civil residente  Sendo sepultado no cemitério de Varges do Ferraço do Servações: La feita a pressura do Servações do S |   |
|             | de sexolución de la companya de de mil novecentos especiales de mortes de la filia de mortes de la filia de mortes de mil novecentos especiales poers de mil novecentos especiales de mil novecentos especiales de militar de mario  |   |
|             | e persenta, per , neste Gartório da premisellado de Municipio de Renge , Estado de Persentalment compareceu fundamente dos facilitàs franctitàs de destro bella franctità dando como causa da morte, presturidade de distribución dando como causa da funcionalmente de funcionalmente de mil novecentos exercita persenta de filmo de con para de filmo de con para de filmo de con profissão Estado Civil residente  Sendo sepultado no cemitério de Varges do Ferraço do Servações: La feita a pressura do Servações do S |   |
|             | c persenta, peis , neste Gartório da premiellado de Municipio de Renge , Estado de Personalment compareceu fundamente dos facilios francei exibindo um estestado de óbito firmado esta dento belso francei dando como causa da morte, presturidade da la comparación de morte, presturidade da la la comparación de mil novecentos esperante personal de mil novecentos esperante personal de sexo properación de con causa de mil novecentos esperante personal de sexo properación de con causa de filo e de follece de filo de constante de filo e de follece de fo |   |
|             | compareceu fundamente dos facilios de compareceu fundamente dos facilios frances exibindo um estestado de óbito firmado colo destro facilio frances dando como causa da morte, fundamente de destro facilio frances dando como causa da morte, fundamenta de decidir fundamente do fundamente do dia funda de fundamente de mil novecentos esperante por la fundamenta de fundamente fundamente de fundamente fundamente de fundamente |   |
|             | compareceu fundamente dos filos de formalidades de apórdo a cir vigente dos fundamentes de de compareceu fundades de como causa da morte, fundade de declarou que que fundamente de dando como causa da morte, fundamente de de mil novecentos escentes pero de fundamente de mil novecentos escentes pero de formal de fundamente de mil novecentos escentes pero de formal de fundamente de mil novecentos escentes pero de formal de formal de fundamente de formal d |   |
|             | compareceu fundamente dos facilios de compareceu fundamente dos facilios frances exibindo um estestado de óbito firmado colo destro facilio frances dando como causa da morte, fundamente de destro facilio frances dando como causa da morte, fundamenta de decidir fundamente do fundamente do dia funda de fundamente de mil novecentos esperante por la fundamenta de fundamente fundamente de fundamente fundamente de fundamente |   |
|             | compareceu fundamente dos filos de formalidades de apórdo a cir vigente dos fundamentes de de compareceu fundades de como causa da morte, fundade de declarou que que fundamente de dando como causa da morte, fundamente de de mil novecentos escentes pero de fundamente de mil novecentos escentes pero de formal de fundamente de mil novecentos escentes pero de formal de fundamente de mil novecentos escentes pero de formal de formal de fundamente de formal d |   |

**Figure 8.** Filtered version of the image in Figure 7.

| Remote full property des de de la comperation de la comparation de la comparation de des de la comparation de de de de la comperation de la comperation de la comperation de la comperation de la competation de l | N.º 19.945. | Aos Juit eties - dias do mês de jaures de mil novecentos   |   |
|--|-------------|--|---|
| Municipio de leure . Etado de le leure . Etado de leure . Etado de le leure . Etado Civil . Etado  | · · /       |  |   |
| exibindo um atestacio de óbito firmado poli dentro feliro Grando dando como causa da morte. Perinaturidade du escrevilidade du escrevilidade du escrevilidade de disconse de mit novecentos e percesto e per la falecia fore filmo en comprando dios de sexo fuerentimo, de côr rando comprando dios de vido natural de ferriparitumo.  Profissão Estado Civil residente  filho de fini, fractario das filmo e de fuerto fuerto de filmo e de fuerto de  |             | Municipio de Reute, Erlado de Ceresculuio -  |   |
| dando como causa da  morte, Servaturidade du  o qual fica arquivado, declarou que ma huaturidade da  liverrillado  dis du horas e du do dia discus ve de  servatura de mil novecentos e persento i para  lateces ori filmo  do sexo monulino, de côr randa compandia dias  du vindo natural de civilario de civil  residente  filho de fina la filmo e de finanta  litario de filmo la filmo e de marta  Observações: La filmo de pois de ter cumprido todas as formalidades de perordo a lei vigente, e depois de tido e achado conjorme, é assinado.  Eu, escrevente juramentafo, escrevi, e cu.  Oficial do Registro Civil, subscrevo.  |             |  |   |
| morte, Pernaturidade du  o qual fica arguivado, declarou que po lucturalidade do  lucirilidado.  dis also horas e des do dia describe de  faleced for horas e des computations de faleced for filmo  do sexo monulino, de côr rondo computations  profissão natural de Pernatural  Profissão Estado Civil  residente  filho de fami leactario de filmo e de functo  fuente da filmo largaderio, monurai  Sendo sepullado no cemiltrio de Varges  Observações: La filmo experimento lodas as formalidades de ogondo a lei vigente, e depois de ler cumprido lodas as formalidades de ogondo a lei vigente, e depois de lido e achado conforme, é assinado.  Eu, cescrevei, c cu, colicial do Registro Civil, subscrevo.   | 1           |  |   |
| o qual fica arguivado, declarou que po haturalade da  licuralizado.  di des — horas e des — do dia deserve — de  laleceu ost felixo — comprais dias  do eccolumentos — de côr randa — comprais dias  de secolumentos — de côr randa — comprais dias  de secolumentos — de côr randa — comprais dias  Profissão — Estado Civil —  residente  filho de fais la etarro da felixo e de huarta  la felixo la felixo la varia —  Sendo sepultado no cemitério de Vargea —  Observações: fo foi la exacura la felixo de la exacura de la exacur |             |  |   |
| de de de proper de mil novecentos e persento, per de la falecea de mil novecentos e persento, per la falecea de mil novecentos e persento de manario de mil novembre de mil novembre de mil novembre de manario de manari |             | morte, Perualui daale di   |   |
| de de de production de con manda de companda de la   |             | o qual lica arquivado, declarou que no Instruidade dos   |   |
| lateces of the and novecentos e persento, are lateces of the and computed dia de sexo months dia de la compute dia de la compute dia de la compute dia la compute de la co |             | Everylliads.   |   |
| falected fort filipo  do sexolumentino, de con manda communication  De sissão Restano de filipo de función  Frofissão Restano de filipo de función  Traine de filipo de función  Sendo sepultado no comitério de la variación  Sendo sepultado no comitério de la variación  Sendo sepultado no comitério de la variación  Deservações: for forte este termo, depois de ter cumprido todas as formalidades de pordo a lei vigente, e depois de lido e achado conforme, e assinado.  Eu, escrevente juramentado, escrevi, e cu,  Oficial do Registro Civil, subscrevo.  |             |  |   |
| do sexo manulino; de côr randa compusati dias  de vida natural de Cumpulano.  Profissão Estado Civil  residente  filho de bui leactano da pluo e de monta  luante da pluo lavadura, naturai,  Sendo sepullado no cemilério de vargos  Observações: La filha examina Dobenico  Decono do Cidado.  Para constar, lavrei este termo, depois de ter cumprido todas as formalidades de apordo a lei vigente, e depois de lido e achado conforme, t assinado.  Eu, escrevente juramentado, escrevi, e eu,  Oficial do Registro Civil, subscrevo.   |             | de mil novecentos e personto per   |   |
| Profissão Estado Civil residente  filho de frui leactare da filoso de tracta  france da filoso de tractar  Sendo sepultado no cemitério de varges  Observações: for for la examina solurio de la estado conforme, te assinado.  Eu, escrevi, e eu,  Oticial do Registre Civil, subscrevo.  |             | falected for files   |   |
| Profissão Estado Civil residente  filho de frui leactare da filvo de morta  Sendo sepultado no cemitério de varges  Observações: for for to a varges  Observações: for for to a varges  Para constar, lavrei este termo, depois de ter cumprido todas as formalidades de otordo a lei vigente, e depois de lido e achado conforme, é assinado.  Eu, escrevi, e eu,  Oticial do Registre Civil, subscrevo.  |             | do sexo Cupullius - de che parda - compuelo dia  |   |
| Profissão Estado Civil residente  filho de bui; leastaur de filos e de marta  Quana de filos lavadura, maturais  Sendo sepultada no cemitério d 1 Varges Observações: for filo a racura soberia p  Para constar, lavrei este termo, depois de ter cumprido todas as formalidades de ceordo a lei vigente, e depois de tido e achado conforme, é assinado.  Eu, escrevente juramentado, escrevi, e eu, Oficial do Registro Civil, subscrevo.  |             | de vido natural de Perupusturo-  |   |
| Para constar, lavrei este termo, depois de ter cumprido todas as formalidades de opordo a lei vigente, e depois de lido e achado conforme, t assinado.  Eu, escrevente juramentado, escrevi, c eu, Oficial do Registro Civil, subscrevo.   |             |  |   |
| Para constar, lavrei este termo, depois de ter cumprido todas as formalidades de peordo a lei vigente, e depois de lido e achado conforme, é assinado.  Eu, escrevente juramentado, escrevi, c eu, Oficial do Registro Civil, subscrevo.   |             | residente  |   |
| Para constar, lavrei este termo, depois de ter cumprido todas as formalidades de peordo a lei vigente, e depois de lido e achado conforme, é assinado.  Eu, escrevente juramentado, escrevi, c eu, Oficial do Registro Civil, subscrevo.   |             |  |   |
| Sendo sepultado no cemitério de la variaca Observações: La fai la secución de la variaca Secución de la la secución de la la secución de la la secución de la  |             |  |   |
| Sendo sepullado no cemilério de la variaca Observações: La fei la servición de la variaca Secuel de la variación de la variación de la variación de la variación de la vigente, e depois de lido e achado conforme, é assinado.  Eu, escrevente juramentado, escrevi, c eu, Oficial do Registro Civil, subscrevo.  |             | I many the second of the secon |   |
| Para constar, lavrei este termo, depois de ter cumprido todas as formalidades de peordo a lei vigente, e depois de lido e achado conforme, é assinado.  Eu, escrevente juramentado, escrevi, c eu,  Oficial do Registro Civil, subscrevo.  | · Alexandra |  |   |
| Para constar, lavrei este termo, depois de ter cumprido todas as formalidades de peordo a lei vigente, e depois de lido e achado conforme, é assinado.  Eu, escrevente juramentado, escrevi, c eu, Oficial do Registro Civil, subscrevo.   |             |  |   |
| dades de opordo a lei vigente, e depois de lido e achado conforme, é assinado.  Eu, escrevente jura- mentado, escrevi, e eu, Oficial do Registro Civil, subscrevo.   |             |  |   |
| dades de opordo a lei vigente, e depois de lido e achado conforme, é assinado.  Eu, escrevente jura- mentado, escrevi, e eu, Oficial do Registro Civil, subscrevo.   |             |  |   |
| dades de opordo a lei vigente, e depois de lido e achado conforme, é assinado.  Eu, escrevente jura- mentado, escrevi, e eu, Oficial do Registro Civil, subscrevo.   |             |  |   |
| dades de opordo a lei vigente, e depois de lido e achado conforme, é assinado.  Eu, escrevente jura- mentado, escrevi, e eu, Oficial do Registro Civil, subscrevo.   |             | Para constar, lavrei este termo, depois de ter cumprido todas as formali-  |   |
| Eu, escrevente jura- mentalo, escrevi, e eu, Oficial do Registro Civil, subscrevo.   |             | dades de perrdo a lei vigente, e depois de lido e achado conforme, & assinado.   |   |
| Oficial do Registro Civil, subscrevo.  |             |  |   |
|  |             | mentalo, escrevi, e eu,  | 1 |
| Julian da Caik   |             | 1 / 1 / /  |   |
|  |             | July on Cant   |   |
|  |             |  |   |
|  |             |  |   |

**Figure 9.** Monochromatic version of Death Certificate after filtering and splitting the image in Figure 8.

The column Thanatos refers to the results obtained in reference [1], while **New** presents the results of the strategy presented in this paper. Table 2 shows that the new strategy presented here presented either no loss or gains in the recognition rate of all fields recognized in relation to the results presented in reference [1]. In the case of the field "Place of death" the increase in recognition rate reached 42%.

#### 4. Conclusions and lines for further work

Handwritten recognition to gain any degree of success either chooses a limited dictionary and allows a large number of writers or widens the vocabulary and largely restricts the numbers of writers. In both cases, the choice of the training set is of central importance for the success of any classification strategy and must be representative of the whole "universe" one wants to correctly recognize. This paper presents a new way of automatically generating the training set for the recognition of a large set of words written by a single user. It has as starting point different sets cursive type fonts, which are modified and compared to the original writing to "match their features". Once the "matching path" is found it is applied to a large dictionary in that encompasses the vocabulary of the document, generating the training set to be used for the whole batch of documents to be transcribed.

The strategy presented here was used with success in two sets of documents. In the case of the transcription of the handwritten letters in the bequest of Joaquim Nabuco it reached the correct rate of 67% transcribed words (of more than three letters), a result that may be considered successful at least for keyword indexing of such historical documents. In the case of death certificates of the Thanatos project, whose vocabulary is far more restricted the results presented either no loss or gains in the recognition rate of all fields recognized in relation to the previous results, reaching an average of 93.79% correct field recognition.

The statistical data collected inter character and inter word spacing, line and character skew, inter line separation were not used to enrich the generation of entries in the dictionary of the training set. Its use is left as a possibility for further work.

#### **Author details**

Gabriel Pereira e Silva and Rafael Dueire Lins *Universidade Federal de Pernambuco, Brazil* 

## Acknowledgement

The authors are grateful to the organizers of the Fontspace site for setting such a useful site, fundamental for the development of this work. The authors also thank the Family Search International Institute for the initiative of digitizing the death certificate records of Pernambuco (Brazil) and to Tribunal de Justiça de Pernambuco (TJPE) to allow the use of such data for research purposes.

Research presented here is partly sponsored by CNPq-Conselho Nacional de Pesquisas e Desenvolvimento Tecnológico, Brazilian Government.

#### 5. References

- [1] A. Almeida, R.D.Lins, and G.F. Pereira e Silva. Thanatos. Automatically Retrieving Information from Death Certificates in Brazil. Proceedings of the 2011 Workshop on Historical Document Imaging and Processing, pp. 146-153, ACM Press, 2011.
- [2] A. I. de S. L. Andrade, C. L. de S. L. Rêgo, T. C. de S. Dantas, Catálogo da Correspondência de Joaquim Nabuco 1903-1906, volume I 1865-1884, volume II 1885-1889, volume III 1890-1910, Editora Massangana, ISBN 857019126X, 1980. (Available at: www.fundaj.gov.br/geral/2010anojn/catalogo\_nabuco\_v2.pdf)
- [3] B. T. Ávila and R. D. Lins. A Fast Orientation and Skew Detection Algorithm for Monochromatic Document Images. 2005 ACM International Conference on Document Engineering, p.118 126. ACM Press, 2005.
- [4] L. Bethell, J. M. De Carvalho. Joaquim Nabuco, British Abolitionists, and the End of Slavery in Brazil: Correspondence 1880-1905, Institute for the Studies of the Americas, 2009. ISBN-13: 978-1900039956.

- [5] M. Bulacu and L. Schomaker. Text-independent writer identification and verfication using textural and allographic features. IEEE Trans. on PAMI, 29(4):701.717, April 2007.
- [6] A. de A. Formiga and R. D. Lins. Efficient Removal of Noisy Borders of Monochromatic Documents. International Conference on Image Analysis and Recognition, 2009, LNCS v.5627. p.158 – 167, Springer Verlag, 2009.
- [7] C. O. A. Freitas, L.S. Oliveira, S.B.K. Aires, F. Bortolozzi, Zoning and metaclasses for character recognition. ACM-SAC 2007. P. 632-636, 2007.
- [8] S. Haykin, Neural Networks: A Comprehensive Foundation. Hardcover (2nd Edition), 1998.
- [9] M. Hu, "Visual pattern recognition by moment invariants", IEEE Transactions on Information Theory, 8(2):179-187, 1962.
- [10] T. Kohonen, Self-Organizing Maps, Springer Series in Information Sciences, Springer, second edition, vol. 30, 1997.
- [11] Z.C. Li, C.Y. Suen, J. Guo, A Regional Decomposition Method for Recognizing Handprinted Characters, IEEE Transactions on Systems, Man, and Cybernetics, N. 25, p. 998-1010, 1995.
- [12] R.D.Lins. Nabuco Two Decades of Document Processing in Latin America, Journal of Universal Computer Science, v. 17(1), pp. 151-161, 2011.
- [13] R.D.Lins, G.F. Pereira e Silva, A.de A. Formiga. HistDoc v. 2.0: enhancing a platform to process historical documents. Proceedings of the 2011 Workshop on Historical Document Imaging and Processing, pp. 169-176, ACM Press, 2011.
- [14] C. Liu, Y. Liu, and R. Dai, "Preprocessing and statistical/structural feature extraction for handwritten numeral recognition", Progress of Handwriting Recognition, A.C. Downton and S. Impedovo eds., World Scientific, 1997.
- [15] C. Liu, K. Nakashima, H. Sako, and H. Fujisawa, "Handwritten digit recognition: benchmarking of state-of-the-art techniques", Pattern Recognition, 36(10):2271-2285, 2003.
- [16] L. Oliveira, R. Sabourin, F. Bortolozzi, and C. Suen, "Automatic recognition of handwritten numerical strings: A recognition and verification strategy", IEEE Trans. on Pattern Analysis and Machine Intelligence, 24(11):1438-1454, 2002.
- [17] G. F. Pereira e Silva and R. D. Lins. An Automatic Method for Enhancing Character Recognition in Degraded Historical Documents. ICDAR 2011, Beijing, September, IEEE Press, 2011.
- [18] C.Y. Suen, J. Guo, Z.C Li, Analysis and Recognition of Alphanumeric Handprints by parts, IEEE Transactions on Systems, Man, and Cybernetics, N. 24, p. 614-631, 1994.
- [19] ABBYY FineReader 12 Professional Editor, http://finereader.abbyy.com/, last visited on 13.04.2012.
- [20] IMAGEJ. http://rsbweb.nih.gov/ij/, last visited on 13.04.2012.
- [21] http://www.fontspace.com/category/cursive?p=19, last visited on 13.04.2012.

- [22] OCRopus 0.3.1 (alpha3): http://code.google.com/p/ocropus/
- [23] Nuance OminiPage Professional 16: http://www.nuance.com/for-individuals/by-product/omnipage/index.htm



