

Document downloaded from:

<http://hdl.handle.net/10251/87627>

This paper must be cited as:

Mauricio Villegas; Toselli ., AH.; Romero Gómez, V.; Vidal, E. (2016). Exploiting Existing Modern Transcripts for Historical Handwritten Text Recognition. IEEE.
doi:10.1109/ICFHR.2016.22.



The final publication is available at

<https://doi.org/10.1109/ICFHR.2016.0025>

Copyright IEEE

Additional Information

© 2016 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Exploiting Existing Modern Transcripts for Historical Handwritten Text Recognition

Mauricio Villegas, Alejandro H. Toselli, Verónica Romero and Enrique Vidal
PRHLT, Universitat Politècnica de València
Camí de Vera s/n, 46022 València, Spain
{mauvilsa, ahector, vromero, evidal}@prhlt.upv.es

Abstract—Existing transcripts for historic manuscripts are a very valuable resource for training models useful for automatic recognition, aided transcription, and/or indexing of the remaining untranscribed parts of these collections. However, these existing transcripts generally exhibit two main problems which hinder their convenience: a) text of the transcripts is seldom aligned with manuscript lines, and b) text often deviate very significantly from what can be seen in the manuscript, either because writing style has been modernized or abbreviations have been expanded, or both. This work presents an analysis of these problems and discusses possible solutions for minimizing human effort needed to adapt existing transcripts in order to render them usable. Empirical results presented show the huge performance gain that can be obtained by adequately adapting the transcripts, thus motivating future development of the proposed solutions.

Keywords-Handwritten Text Recognition; Historical Manuscripts; Modernized Transcripts; Transcript-image Alignment, Diplomization

I. INTRODUCTION

From works of historians and paleographers there are many historical handwritten collections that have been partially transcribed. These existing transcripts, are a very valuable resource for training models that can be used for automatic handwriting text recognition (HTR), or indexing by means of keyword spotting (KWS), of the remaining untranscribed parts of these collections.

However, since these transcripts were generated for different purposes, its use for HTR or KWS entails some challenges. One problem is the alignment of manuscript and transcript text: paging information from the original document is generally missing, transcriber notes are often added to the actual transcripts, parts of the manuscript are left untranscribed, etc. A different kind of problem comes from the changes that transcribers have introduced with respect to what was actually written in the original manuscript. Examples of types of changes are: expansion of abbreviations, joining of words broken between lines, modified or added punctuation, added accents, modernization of word spellings and correction of writing mistakes.

Transcripts which have been produced with these types of modifications will be called “*modernized*” transcripts throughout this paper. Clearly, these transcripts lack a sequential, one-to-one correspondence with character image segments, which make them inadequate for their use in

segmentation-free HTR training techniques. For adequate training HTR models, the transcripts should ideally correspond exactly to the actual characters as observed in the manuscript. These transcripts, generally referred to as “*diplomatic*” transcript, are very scarce because they are often difficult to read and are generally useful only for paleography studies. In addition, they are considerably more expensive to produce.

This paper presents a study of the challenges encountered when using modernized transcripts in HTR, discusses how they can be profitably used, and presents experimental results that give an indication of the benefits that can be obtained. In [Sec. II](#) the problem is stated, analyzed and research directions are proposed for dealing with the different challenges. It does not attempt to propose a single approach for adapting a transcript, but discuss several possibilities considering the various amounts of manual work required. Then [Sec. III](#) describes a dataset for which ground truth was created in order to understand better and quantify the problems on real data, and to provide data for the experiments that are presented in [Sec. IV](#). The final section concludes the paper, outlining the future directions of research that should be developed to take full advantage of existing transcripts.

II. PROBLEM STATEMENT

The scenario considered in this paper is one in which there is a large collection of more or less homogeneous manuscript images, i.e., all the images share some common characteristics such as being from the same author, written by the same hands, or being of the same period and topic. Also, it is assumed that part of this collection has been transcribed, however, these transcripts are *modernized*; i.e., not identical or aligned with the original manuscript (c.f., [Sec. I](#)). In most cases, it is desired to transcribe the complete collection, or to index it so that it becomes searchable, thus the use of automatic or semi-automatic HTR technologies is in order. Furthermore, due to the homogeneity of the collection, training specific models for this data is expected to give significantly better results than using more generic systems or models.

The principal question to tackle is how to obtain the best recognition or indexing performance by gaining leverage from existing transcripts, while requiring the least human effort to make use of these transcripts. Obviously, the

answer highly depends on certain aspects of the transcripts available, such as the information they contain, in which format, and the amount and kind of modernization of these transcripts. Two main, non-exclusive challenges are identified: the manuscript-transcript alignment and the differences between the text in the manuscript images and that in the modernized transcripts.

A. Aligning Manuscript Images and Transcripts

Currently, the most popular models for HTR are Hidden Markov Models (HMMs) trained using either the Baum-Welch or the Viterbi algorithms, and Recurrent Neural Networks (RNNs) trained using Connectionist Temporal Classification (CTC). Both of these require segmented image text lines and the corresponding transcript for each of the lines. Thus, to make use of an existing transcript, the first step is to obtain a per manuscript line transcript alignment. Luckily it is common that transcripts include some kind of alignment information, i.e. where the corresponding text appears in the original manuscript. Obviously, for training HTR models, the most detailed this information, the better. As commented above, HMM or RNN training typically requires text-to-image alignment at the line level. Though, most commonly only page level text-to-image alignment are available (or can be cost-effectively produced manually); that is, which chunks of the text of the given transcript correspond to each of the manuscript pages. Interestingly, this rough information can be very useful in cases in which the reading order of the page is clearly defined, since a relatively good transcript line segmentation can be obtained by doing a forced alignment at a page level [1], [2].

A clear reading order, however, cannot always be assumed. Manuscript pages may have a complex layout and various text blocks with arbitrary reading orders. There can be parts of the manuscript that are not transcribed, thus text in the images should be aligned to nothing. In some cases these are very short, such as page numbering, and in other cases can be longer, for example skipped side notes. On the other hand, if the manuscript text lines are automatically detected then dealing with the detection errors require even more complex line level text-to-image alignments where parts of the transcript should be left unaligned, even though they do appear in the manuscript. Another difficulty appears when there are text fragments in the transcript which do not appear in the manuscript, for example transcription notes.

Due to this wide variability in the type of given transcripts, as well as in the kind of manuscript location information implicitly or explicitly available in these transcripts, line-level text-to-image alignments have to be produced manually in many cases. Therefore, semi-automatic methodologies are needed to aid in this process, requiring the least human effort possible. The hardest case would be a transcript without any location information, having arbitrary reading order text blocks and including transcription notes. To lower the burden

to the user, it would be desirable to have a system that given the whole transcript, assigns text to pages and within page regions. This assignment would then be checked by a user to correct region level errors and to confirm which parts are not manuscript text. One possibility to achieve this could be to do a recognition of the manuscript images and use the result to do the page/region level alignment. However, available recognition models could give useless results for a given manuscript. Nevertheless, research in this direction should be conducted.

One final detail to comment about the alignment problem is that if the transcript is not diplomatic, as assumed in this work, then the existing alignment techniques cited above will tend to fail. To overcome this, new alignment methods are needed which simultaneously consider not only the text-to-manuscript alignment problem, but also the task of diplomatization of the transcript, discussed in the following subsection.

B. Diplomatization of Transcripts

As mentioned before, to adequately train HTR models, diplomatic transcripts are required, but these are generally unavailable. There are many types of differences between these transcripts and the modernized transcripts typically available:

- Expansion of abbreviated words,
- Modified or added punctuation,
- Modernization of word spellings,
- Joining inter-line broken words and removal of word-break marks (e.g. hyphens),
- Correction of writing mistakes in the manuscript,
- Typing mistakes in the transcripts,
- Missed or duplicate words in the transcripts.

Manually undoing all of these changes is considerably expensive, especially for historical manuscripts that normally require expert paleographers. Therefore, developing techniques that automatically or semi-automatically perform “diplomatization” of transcripts is needed. When doing this diplomatization it is important to keep the correspondence between each diplomatic word form and its modern version, since this allows HTR systems to recognize words as diplomatic-modernized pairs, thereby allowing to produce both diplomatic and modernized transcripts. Depending on the application, this additional information can be very valuable; for example in indexing, it would make it possible to search for abbreviated words by issuing a query either as its abbreviated or the expanded form.

The diplomatization problem is encountered in diverse corpora. In the Alcaraz dataset (see [Sec. III](#)) for about 26% of the words, the modernized version differs from the diplomatic one. A similar trend can be observed in other collections: 22% in both the Wiensankturlrich [3] and the Saint Gall [2] datasets, and 28% in the HATTEM dataset [4]. The amount of abbreviations may depend a

lot on the language and the document, and abbreviations are not the only change possible. So the discrepancies can be much higher in some cases. An example is the large French-Latin collection called “Chancery”¹, encompassing about 25,000 densely handwritten page images, which is being considered in the HIMANIS project². Roughly half of the text is handwritten in medieval French and the other half in Latin. In this case, it has been estimated that about 60% of the Latin words and 20% of the French words are abbreviated. Parts of “Chancery” manuscripts have modern transcripts available, estimated to be around half a million words, and in all of these the abbreviations are expanded.

In some cases, abbreviation and spelling modernization tables are available, which can be used to attempt automatic diplomatization to some extent. However, this is not available for every language and time period and, when they are, the information is generally highly ambiguous; that is, a given modernized word corresponds to many paleographic renderings and a paleographic word form admits many modernized versions.

Nevertheless, compiling these tables often requires relatively little effort, in comparison to manually diplomatizing a document of moderate size. So efficient diplomatization should start by first creating tables of diplomatic-modern mappings. Then, a system could be built that automatically diplomatizes a modernized transcript. However, given the ambiguities commented above, this approach is problematic since it often results in huge amounts of alternative diplomatized versions of a single line of a modernized transcript. A possible solution would be a system that takes as input both a transcript and a manuscript image, and starts suggesting diplomatic-modernized mappings, possibly with accompanying example images. After a user validates a part of the diplomatic-modernized mappings, the system could present the automatically diplomatized manuscript in a graphical interface for the user to interactively correct it, a procedure very similar to the one used in interactive transcription systems such as CATTI [5], [6].

III. ALCARAZ DATASET

As a case study of the problems discussed in Sec. II, the dataset chosen is a subset of the written records from the Inquisition process (1534-1539) against Pedro Ruiz de Alcaraz, a member of the Spanish *alumbrado* religious movement in the 16th century. See Fig. 1 for sample pages. The complete dataset is composed of 953 page images written in Spanish using mainly two types of calligraphies known as *Cortesana* and *Procesal*. Since these are records of a trial, they are characterized by being quick on-line writing, without consistent blank spaces between and within words, and with plenty of (rather inconsistent, often improvised) abbreviations. Partial modernized transcripts were produced for

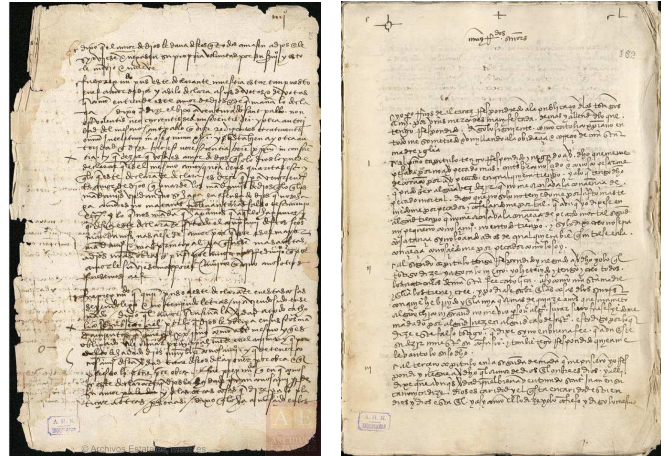
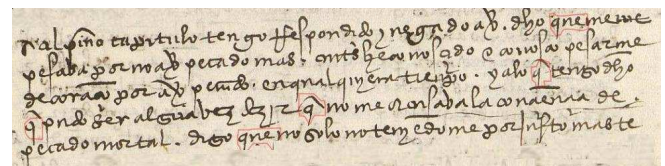


Figure 1. Example of page images from the Alcaraz dataset.



Diplomatic transcript (needed, but unavailable):

§ *al pmo capitulo* tengo respondido y negado *avr dho que* me *me* pesava por no *avr* pecado *mas* . *ants* he conocido y conosco pesarme de *coraçon* por *avr* pecado en qualquiera tiempo . y a lo *q* tengo *dho q* *pud* ser alguna vez *dzir q* no me acusava la conciencia de pecado mortal . *digo que* no *solo* no *teniedome* por justo mas *te*

Aligned modernized transcript

ten. Al primero capitulo tengo respondido y negado *aver dicho que me* pesava por no *aver* pecado *mas* . *Antes* he conocido y conosco pesarme de *coraçon* por *aver* pecado en qualquiera tiempo . Y a lo *que* tengo *dicho que pudo* ser alguna vez *dezir que* no me acusava la conciencia de pecado mortal . *Digo que* no *sólo* no *teniéndome* por justo, mas *teniéndome*

Original modernized transcript (partially available):

Iten. Al primero capítulo tengo respondido y negado aver dicho que me pesava por no aver pecado más. Antes he conocido y conosco pesarme de coraçon por aver pecado en qualquiera tiempo. Y a lo que tengo dicho que pudo ser alguna vez dezir que no me acusava la conciencia de pecado mortal. Digo que no sólo no teniéndome por justo, mas teniéndome [...]

Figure 2. Excerpt from the Alcaraz dataset showing the original image and its diplomatic and modernized transcripts. The words in blue-italic font are different in the diplomatic and the aligned modernized versions. The appearances of the word “que” are marked in the image surrounded by a red polygon and in the transcripts in red-bold font. This is an example of modernized word that may or may not appear abbreviated in the image – and therefore in the diplomatic transcript.

this manuscript by Kinder [7], but no diplomatic transcripts are available. Fig. 2 shows an image fragment of this dataset, along with the initially unavailable aligned and paleographic transcripts and the available modernized transcript. The modernization changes include: added capitalization and punctuation, spelling standardization, changed, removed or added words, and expanded abbreviations.

From the complete dataset, a subset of 44 manuscript pages was chosen. These pages correspond to some of the declarations made by Pedro Ruiz de Alcaraz and are written in *Cortesana* calligraphy. Modernized transcripts from [7] are available for these pages. These 44 pages and their available transcripts were processed to obtain adequate ground

¹<http://www.culture.gouv.fr/documentation/archim/tresor-chartes.html>

²<http://himanis.hypotheses.org>

Table I
STATISTICS FOR THE 44 PAGE SUBSET OF THE ALCARAZ DATASET.

Number of:	Absolute	Relative/Average
Pages	44	
Lines	1,728	39.3 per page
Broken diplomatic words	126	7.3 every 100 lines
Running diplomatic words*	23,481	13.6 per line
Running modernized words	24,106	14.0 per line
Running diplomatic characters [†]	111,911	3.8 per dipl. word
Running modernized characters [†]	124,233	4.3 per dipl. word
Words diplomatic \neq modernized [‡]	6,162	26.1% of runn./words
Lexicon size	diplomatic=3,235	modernized=2,669
Character set size [§]	diplomatic=31	modernized=30

* Broken words counted as one.

[†] Absolute includes between-word spaces.

[‡] Capitalization insensitive comparison.

[§] Manuscript does not distinguish lower/upper-case and has no accents.

truth for analyzing the modernized transcript problem being considered. First the text lines of the manuscript were automatically detected, followed by a manual correction of the baselines. In parallel, the transcripts were manually aligned with the corresponding manuscript page images. Then, a forced alignment process was performed as explained in Sec. IV-A, resulting in a segmentation of the transcript text into the lines of the manuscript. Finally, the aligned transcript was manually checked and augmented, a task that included: possible correction of the previous automatic text segmentation errors, marking and splitting of the words broken between lines, adding of untranscribed lines and words, typo corrections, and adding to each word the corresponding diplomatic version if different from the modernized one. Statistics of the 44 page dataset are presented in Tab. I.

IV. EXPERIMENTAL EVALUATION

The results of two experiments are presented in this work, each one related to the two challenges discussed in Sec. II.

A. Page-based Forced Alignment

This experiment shows the alignment performance when having as input manuscript images with detected and checked lines, and each page having the corresponding modernized transcript. The objective is to assign the words to the manuscript lines. The alignment was done using HMMs with the technique presented in [1], with the addition that words aligned to transitions between lines are assumed to be broken words. The pre-processing included: extraction of line images, correction of line slope and slant, noise removal and image enhancement [8]. Feature vectors were extracted using the technique from [9].

Since initially there were no models appropriate for this data, HMMs were trained from scratch using the modernized transcripts and all 44 pages. In order to do this, the feature vectors of all the lines in a page were concatenated, and this along with the modern page transcript was used both for training HMMs and Viterbi forced alignment. The HMMs

Table II
RESULTS FOR THE PAGE ALIGNMENT EXPERIMENT.

Complete lines correctly aligned	1,289	74.6% of lines
Words correctly aligned	22,875	97.4% of words
Words aligned to incorrect line	599	2.6% of words
Correctly identified broken words	84	66.7% of broken words
Words incorrectly broken	119	6.9 every 100 lines

had 6 states and Gaussian Mixture Models (GMM) of 16 components and were trained for 5 Expectation Maximization (EM) iterations using HTK³.

The results are presented in Tab. II. The counts are with respect to diplomatic words, since this is what defines a correct line in the manuscript. The alignment is quite good, 97.4% of words are assigned correctly. The breaking of words does not seem to be useful since there are more false positives than broken words correctly identified. Even though these alignment results are impressive and HMM models are obtained as a byproduct, they provide poor recognition performance as can be observed next.

B. Diplomatic-Modernized HTR Performance Comparison

This experiment shows the performance gain of using a perfect diplomatic transcript in comparison to having only the modernized transcript. It is intended to motivate the need to develop systems for diplomatization. Three different HTR usage scenarios are considered, namely: automatic handwritten text recognition, computer assisted transcription, and keyword spotting. For each scenario results are presented for two extremes: 1) Modernized training and 2) Diplomatic training. The latter includes diplomatic-expansion pairs, but in decoding only the expansion is kept so that it is comparable to modernized training.

All results are based on a 10-fold cross-validation experiment, where each of the 10 partitions corresponds to 4 or 5 of the manuscript pages. For each partition both the character models and the language model were trained using only the other 9 partitions, i.e., no external data was used.

We used a classical HTR system architecture composed of three modules: image pre-processing, line image feature extraction and HMM and language model training/decoding [10]. The pre-processing was the same as in Sec. IV-A. Training of HMMs and decoding was done using HTK³ and bi-gram language models trained using SRLIM⁴. The HMM character models had 6 states, GMMs of 64 components and were trained for 8 EM iterations. The HVite decoder was used to generate word graphs (WG) using input degree of 15, Grammar Scale Factor (GSF) of 10 and Word Insertion Penalty (WIP) of 0.

Automatic Handwritten Text Recognition: In this HTR usage scenario the goal is to obtain a recognized transcript without any human intervention. As previously commented,

³<http://htk.eng.cam.ac.uk>

⁴<http://www.speech.sri.com/projects/srlim/>

Table III
RESULTS (IN %) FOR THE THREE HTR USAGE SCENARIOS.

	Automatic Recognition		Assisted Transcription		Keyword Spotting	
	WER	CER	WSR	EFR	AP	mAP
Modernized training	58.1	29.9	40.1	30.9	64.6	70.8
Diplomatic training	43.4	19.6	24.2	44.2	77.8	84.4
Relative improvement	25.3	34.4	39.7	19.2	37.3	46.6
Out-of-vocabulary (OOV) words modern:						23.7%
Running out-of-vocabulary (ROOV) words modern:						7.3%

for each test line, a WG was generated. Using these WGs, rescoring was performed to get the 1-best recognition, trying all combinations of $GSF=\{0, 3, 5, 10, 20, 30, 50\}$ and $WIP=\{50, 30, 20, 10, 5, 0, -5, -10, -20, -30, -50, -70, -90, -110\}$. The best parameters both for modern and diplomatic training were $GSF=10$ and $WIP=20$.

The quality of the recognition is given by the well known *word* and *character error rates* (WER and CER). They are defined as the minimum number substitutions, deletions and insertions of words/characters need to convert the text produced by the system into the reference transcripts, divided by the total number of words/characters in the reference.

The results are presented in the first two columns of Tab. III. For both metrics, WER and CER, the performance is significantly better for diplomatic training than modernized, having a relative improvements of 25% and 34% respectively. The error rates are relatively high, however, this is just an indication of the difficulty of this dataset.

Computer Assisted Transcription: In the computer assisted transcription scenario the user and the HTR system work jointly, to obtain perfect transcripts of the text images. The human transcriber is directly involved in the transcription process by being responsible of validating and/or correcting the HTR output as it is being produced. The goal is to obtain the perfect transcript with the least human effort possible. The approach used here is presented in detail in [6] and is called “Computer Assisted Transcription of Text Images” (CATTI). The results presented here are based on the generated WGs, since this is what CATTI requires.

To assess the performance we use the *word stroke ratio* (WSR) and the *estimated effort reduction* (EFR). WSR is defined as the number of word level user interactions necessary to get the reference transcript, divided by the total number of words in the reference. It gives an estimate of the (simulated) human effort needed to produce correct transcripts. The EFR is the relative difference between WER and WSR that provides an estimate of the reduction in human effort using CATTI with respect to manually correcting the output of automatic HTR.

Results are presented in the middle two columns

of Tab. III. According to these, to produce 100 words of correct transcripts, a user would have to type 40 words (the remaining are predicted automatically) if CATTI with modernized training is used. However, for diplomatic training the number of typed words goes down to 24. A similar trend is observed for EFR, with modern training 31% of the errors would be automatically corrected by the interactions in comparison to a much higher 44% for diplomatic.

Keyword Spotting: The indexing and search approach presented here is *line-based*, that is, the search domain is a set of text line images. The goal is to determine whether a given keyword appears in each text line, regardless of number of occurrences. The approach adopted here is the one presented in [11], taking as input the WGs normalized according to the procedure described in [12].

For effectiveness assessment of the KWS approach, the standard *recall* and *interpolated precision* measures [13] are used. We employ another popular scalar KWS assessment measure called *average precision* (AP) [14] which, actually, is the area under the *recall-precision* curve. In addition, the *mean average precision* (mAP) is also used, which is very often adopted in the KWS literature. It is computed by averaging the *average precision* of each keyword.

Fig. 3 shows the *recall-precision* curves and the last two columns of Tab. III show the AP and mAP figures. These correspond to averages over the 10 cross-validation partitions. To compute the mAP, only the words found in each partition test set are considered, since mAP is ill-defined when there are zero relevant samples. For the AP and recall-precision curves, all the words in the training set are considered.

As observed in Fig. 3, for low values of recall the precision is very similar for both training types. However, for higher values of recall, diplomatic training is significantly better. This behavior is mainly due to modernized words which are very different from their diplomatic counterparts, leading to poor training for some of the characters. An example of this fact occurs with the (modernized) word “*Iten.*”, which can only be (inadequately) modeled with all its five characters in modernized training, while in diplomatic training it is modeled with its only correct character “*§*”, as it appears in the image. The resulting differences in training quality affect directly the word confidence scores, which tend to be very low with modernized training and very high for diplomatic training. To retrieve all diplomated variants of a modernized word, modeling correctly the characters as they appear in the manuscript is fundamental.

V. CONCLUSIONS AND FUTURE WORK

This paper deals with the problem of profiting from existing modernized transcripts, by adapting them with the least effort possible, so that they are better suited for training handwritten text recognition models. For the challenge of aligning transcript text to the manuscript lines, current forced

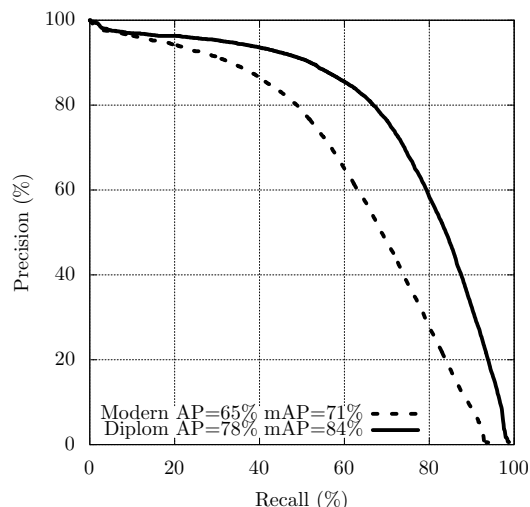


Figure 3. KWS Recall-Precision curve (with interpolated precision) plotted for trained HTR with diplomatic and modern transcripts respectively.

alignment techniques perform very well despite the discrepancy of the modernized text. However, transcripts need to be provided for each manuscript page, there can't be much extra or missing text and the reading order must coincide. To reduce manual effort, future works should attempt alignment at a multi-page level and be able to cope with possible unmatched text and variations in reading order.

Regarding the other main challenge discussed, empirical results presented show that by diplomatising the transcripts, huge improvements in performance are obtained for three usage scenarios considered: recognition, assisted transcription and search. This highly motivates future developments of proposed ideas, in particular: techniques for suggesting possible diplomatic-modernized mappings that would be user validated and, based on these mappings doing automatic diplomatisation that could then be interactively corrected and validated in a manner similar to current assisted transcription systems.

ACKNOWLEDGMENTS

We are very grateful to Carlos Lechner and Celio Hernández who helped in the creation of the ground truth of the Alcaraz dataset. This work has been partially supported by the European Union (EU) Horizon 2020 grant READ (Recognition and Enrichment of Archival Documents) (Ref: 674943), EU project HIMANIS (JPICH programme, Spanish grant Ref: PCIN-2015-068) and MINECO/FEDER, UE under project TIN2015-70924-C2-1-R.

REFERENCES

[1] A. H. Toselli, V. Romero, and E. Vidal, "Alignment between text images and their transcripts for handwritten documents," in *Language Technology for Cultural Heritage*. Springer, 2011, pp. 23–37, doi:10.1007/978-3-642-20227-8_2.

[2] A. Fischer, V. Frinken, A. Fornés, and H. Bunke, "Transcription alignment of latin manuscripts using hidden markov models," in *Proceedings of the 2011 Workshop on Historical*

Document Imaging and Processing. ACM, 2011, pp. 29–36, doi:10.1145/2037342.2037348.

[3] V. Romero, A. Toselli, J. Sánchez, and E. Vidal, "Handwriting transcription and keyword spotting in historical daily records documents," in *International Workshop on Document Analysis Systems (DAS)*, Santorini, Greece, April 2016, pp. 275–280.

[4] J. A. Sánchez, V. Bosch, V. Romero, K. Depuydt, and J. de Does, "Handwritten text recognition for historical documents in the transcriptorium project," in *Proceedings of the First International Conference on Digital Access to Textual Cultural Heritage*, ser. DATeCH '14. New York, NY, USA: ACM, 2014, pp. 111–117, doi:10.1145/2595188.2595193.

[5] A. Toselli, V. Romero, M. Pastor-i-Gadea, and E. Vidal, "Multimodal Interactive Transcription of Text Images," *Pattern Recognition*, vol. 43, no. 5, pp. 1814–1825, 2010.

[6] V. Romero, A. H. Toselli, and E. Vidal, *Multimodal Interactive Handwritten Text Transcription*, ser. Series in Machine Perception and Artificial Intelligence (MPAI). World Scientific Publishing, 2012.

[7] A. G. Kinder, "Inquisition documents concerning Pedro Ruiz de Alcaraz," in *The Kinder Collection in the John Rylands University Library*, ser. Special Collections, <http://www.library.manchester.ac.uk/search-resources/guide-to-special-collections/atoz/kinder-collection/>.

[8] M. Villegas, V. Romero, and J. A. Sánchez, "On the Modification of Binarization Algorithms to Retain Grayscale Information for Handwritten Text Recognition," in *7th Iberian Conference on Pattern Recognition and Image Analysis*, ser. LNCS. Santiago de Compostela (Spain): Springer, Jun. 2015, vol. 9117, pp. 208–215, doi:10.1007/978-3-319-19390-8_24.

[9] M. Kozielski, J. Forster, and H. Ney, "Moment-based image normalization for handwritten text recognition," in *Frontiers in Handwriting Recognition (ICFHR), 2012 International Conference on*, Sept 2012, pp. 256–261, doi:10.1109/ICFHR.2012.236.

[10] A. H. Toselli, A. Juan, D. Keysers, J. González, I. Salvador, H. Ney, E. Vidal, and F. Casacuberta, "Integrated Handwriting Recognition and Interpretation using Finite-State Models," *Int. Journal of Pattern Recog. and Artificial Intelligence*, vol. 18, no. 4, pp. 519–539, 2004, doi:10.1142/S0218001404003344.

[11] A. H. Toselli, E. Vidal, V. Romero, and V. Frinken, "Word-graph based keyword spotting and indexing of handwritten document images," Universitat Politècnica de València, Tech. Rep., 2013.

[12] —, "Word-graph based keyword spotting in handwritten document images," 2013, under review.

[13] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. New York, NY, USA: Cambridge University Press, 2008.

[14] S. Robertson, "A new interpretation of average precision," in *Proc. of the International ACM SIGIR conference on Research and development in information retrieval (SIGIR '08)*. New York, NY, USA: ACM, 2008, pp. 689–690, doi:10.1145/1390334.1390453.