Recognition and Enrichment of Archival Documents

D6.10.
Line and Word Segmentation Tools P1

Georgios Louloudis, Nikolaos Stamatopoulos, Basilis Gatos, NCSR Demokritos

Distribution:

http://read.transkribus.eu/

READ
H2020 Project 674943

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 674943

Project ref no.	H2020 674943
Project acronym	**READ**
Project full title	**Recognition and Enrichment of Archival Documents**
Instrument	H2020-EINFRA-2015-1
Thematic Priority	EINFRA-9-2015 - e-Infrastructures for virtual research environments (VRE)
Start date / duration	01 January 2016 / 42 Months

Distribution	Public
Contractual date of delivery	31/12/2016
Actual date of delivery	28/12/2016
Date of last update	21/12/2016
Deliverable number	D6.10
Deliverable title	Line and Word Segmentation Tools P1
Type	Demonstrator
Status & version	Public & version 1
Contributing WP(s)	WP6
Responsible beneficiary	NCSR
Other contributors	CVL, UPVLC
Internal reviewers	UPVLC, EPFL
Author(s)	Georgios Louloudis, Nikolaos Stamatopoulos, Basilis Gatos NCSR UPVLC CVL
EC project officer	Martin Majek
Keywords	Text Line Segmentation, Word Segmentation

Table of Contents

Executive Summary

This deliverable reports on the achievements concerning the tasks of text line and word segmentation at the end of the first year of the READ project. The NCSR group has a strong experience on these tasks, organized several related competitions and was involved in the previous "IMPACT" and "tranScriptorium" EU projects that processed machine-printed and handwritten historical documents, respectively. Based on this experience, we can state that problems and challenges become significantly stronger when going from machine-printed to handwritten documents which are the main focus of the "READ" project.

1. Text Line Segmentation

One of the early tasks in a handwriting recognition system is the segmentation of a handwritten document image into text lines, which is defined as the process of defining the region of every text line on a document image. It should be stressed that the expected input to this module is a single column text region which is actually the output of the basic layout analysis module (task 6.2). To this end, the effectiveness of the text line segmentation process is strongly related with the result of the layout analysis stage. At the same time, results of poor quality produced by the text line segmentation stage seriously affect the accuracy of the handwritten text recognition procedure. Several challenges exist on historical documents which should be addressed by a text line segmentation method. These challenges include: a) the difference in the skew angle between lines on the page or even along the same text line, b) overlapping and touching text lines, c) additions above the text line and d) deleted text. Figure 1.1 presents one example for each of these challenges. It should also be stressed that two main ways exist for representing the results of a text line segmentation method: i) using polygons that enclose the corresponding text lines and ii) using baselines i.e. a set of (poly)line segments which correspond to the imaginary lines on which the scribe writes the text. Figure 1.2 presents one example of each of the abovementioned representation ways.

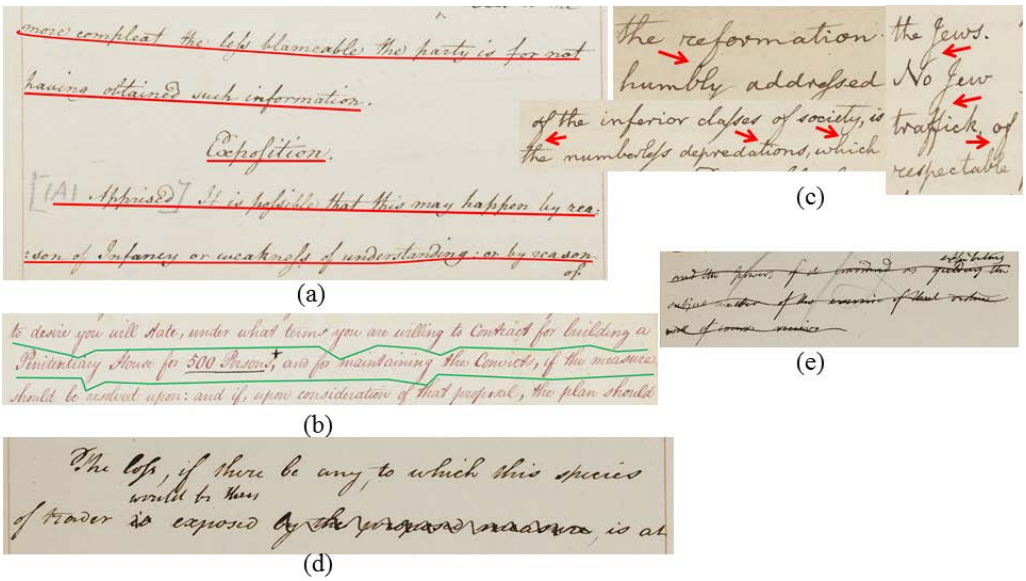


Figure 1.1: Challenges encountered on historical document images for text line segmentation: (a) Difference in the skew angle between lines on the page or even along the same text line, (b) overlapping text lines, (c) touching text lines, (d) additions above a text line, e) deleted text.

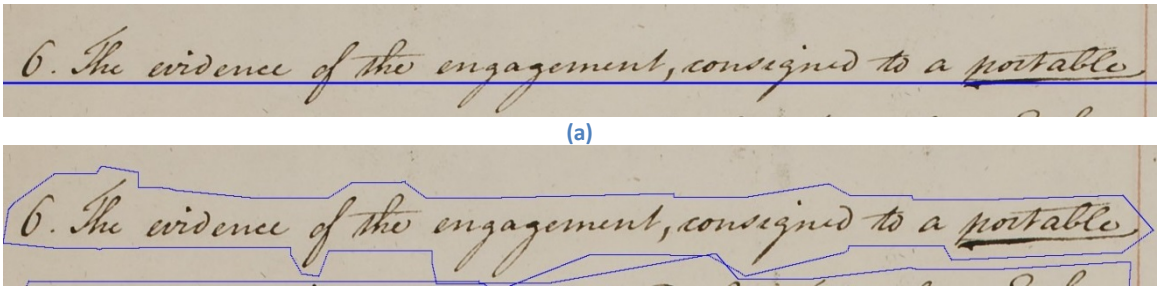


Figure 1.2: Representation of the text line segmentation result using (a) baseline and (b) polygon.

1.1. NCSR Text line Segmentation Method – 1st Year

One of the outcomes of the "tranScriptorium" project was the development of the NCSR text line segmentation method -NCSR (1st year) method -able to deal efficiently with most of the challenges encountered in this field of research. It was an extension of the methods presented in [Louloudis2009] as well as in [Gatos2014]. More specifically, the existing method was adapted to the nature and characteristics of historical handwritten documents. The developed method contains two differences with respect to the references mentioned above. The first difference concerns the development of a baseline estimation method starting from the polygon based representation. The reason for that was the enormous effort needed to correct the erroneous regions encountered using the polygon representation. The introduction of baselines made the correction quicker and more efficient. In more detail, the developed method defines a baseline as a set of points whose number depends on the size of the text line. The position of the lowest black pixel is computed for each column of the text line image and all these points build the set upon which the method is working. At a next step, a linear regression on this set of points is applied. For the case of small text lines, the regression is applied on the whole text line image. Large (with respect to their width) text lines are split into three uniform segments for which a separate regression is applied. Figure 1.1.1 presents examples of the produced baselines on a small text line (a), on a large text line (b) and on a large fluctuating text line (c).

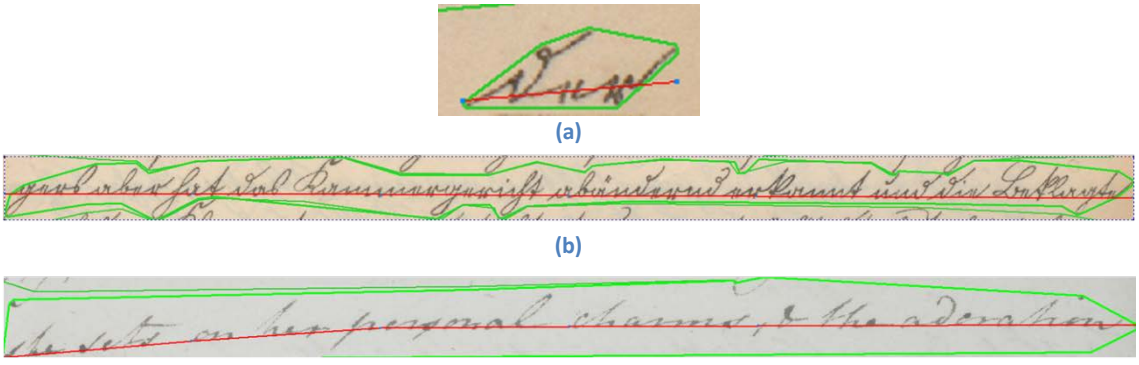


Figure 1.1.1: Examples of the produced baselines on a (a) small, (b) large and (c) fluctuating text line.

The second difference compared to the already published NCSR methods concerns the extension of the polygon based method by (a) making use of the extracted baselines described above and (b) adding a fourth stage which tries to solve the majority of the errors

encountered by the initial method. More details concerning the developed method can be found at [Gatos2015].

In addition to the method described above, we also developed a text line segmentation method which produces the polygon representation when the input corresponds to the ground truth baselines (FromBaseToPoly method). The method assigns the connected components of the document image to the "closest" text line. For the case of touching components, an intelligent method is applied making use of the components skeleton in order to efficiently assign the pixels of the component to the correct text line. The final step of the procedure concerns the creation of the polygon representation starting from the pixel based representation. The polygon creation is based on an efficient algorithm which creates text line polygons with a small set of vertices [Retsinas2016].

It should be stressed that the NCSR text line segmentation as well as FromBaseToPoly methods are developed in C++ following the guidelines of the Transkribus interface and are available at github:

https://github.com/Transkribus/NCSR_Tools

1.2. Evaluation Protocol

In order to measure the performance of the NCSR text line segmentation method (1st year), two different evaluation protocols were used: a) polygon based evaluation, b) baseline evaluation.

For the polygon based evaluation we followed the same protocol which was used in the ICDAR 2013 Handwriting Segmentation Competition [Stamatopoulos2013]. According to this protocol, the performance evaluation is based on counting the number of one-to-one matches between the areas detected by the algorithm and the areas in the ground truth (manual annotation of correct text lines).

We consider a region pair as a one-to-one match only if the matching score is equal to or above the evaluator's acceptance threshold T_a. If N is the count of ground-truth elements, M is the count of result elements, and *o2o* is the number of one-to-one matches, we calculate the **polygon Recall** (*PR*) and **polygon Precision** (*PP*) as follows:

$$PR = \frac{o2o}{N} \quad (1.2.1) \qquad PP = \frac{o2o}{M} \quad (1.2.2)$$

A performance metric **polygon F-Measure** (*PFM)* can be extracted if we combine the values of polygon Recall and polygon Precision:

$$PFM = \frac{2 * PR * PP}{PR + PP} \quad (1.2.3)$$

For the baseline evaluation protocol, a modification of the precision, recall, F-measure approach described above is considered. In more detail, to define precision and recall we need some kind of "counting" function which counts the number of points of *p* for which there is a point of *q* with a distance less than *t*. The modified metrics, for which we will use the names **baseline Recall** (*BR*), **baseline Precision** (*BP*) and **baseline F-Measure** (BFM) in order to distinguish them from the polygon based metrics are explained in detail in [Gruning2016].

1.3. Experimental Results

The performance of the NCSR text line segmentation method (1st year), the state-of-the-art method presented in [Louloudis2009] as well as the method which produces polygons when correct baselines are used as input (FromBaseToPoly method) have been tested using three challenging datasets of historical handwritten documents: (i) Konzilsprotokolle (German), (ii) NAF (Finnish) and (iii) BL (English). Table 1.3.1 summarizes the number of documents together with the number of text lines and words for each dataset. An analytical description of these datasets can be found in deliverable D7.13 "Keyword Spotting Engines QbS, QbE".

Table 1.3.1: Summary of dataset information used to evaluate the text line and word segmentation methods.

Dataset	#documents	#text lines	#words
Konzilsprotokolle (German)	100	2555	15567
NAF (Finnish)	56 (double pages)	3186	16201
BL (English)	115	2971	15739

Tables 1.3.2 – 1.3.4 present comparative experimental results for each dataset using the polygon based evaluation in terms of PR, PP, and PFM. The acceptance threshold (T_a) was set to 0.95.

Table 1.3.2: Comparative experimental results using Konzilsprotokolle dataset (polygon evaluation)

Method	# GT lines	# RS lines	#o2o	PR	PP	PFM
NCSR (1st year)	2555	2532	2289	89.59	90.40	90.00
FromBaseToPoly	2555	2555	2454	96.05	96.05	96.05
Louloudis2009	2555	2520	2178	86.43	85.24	85.83

Table 1.3.3: Comparative experimental results using NAF dataset (polygon evaluation)

Method	# GT lines	# RS lines	#o2o	PR	PP	PFM
NCSR (1st year)	3186	3053	2756	86.50	90.27	88.35
FromBaseToPoly	3186	3177	2976	93.41	93.67	93.54
Louloudis2009	3186	3085	2725	85.53	88.33	86.91

Table 1.3.4: Comparative experimental results using BL dataset (polygon evaluation)

Method	# GT lines	# RS lines	#o2o	PR	PP	PFM
NCSR (1st year)	2971	2888	2282	76.81	79.02	77.90
FromBaseToPoly	2971	2969	2847	95.83	95.89	95.86
Louloudis2009	2971	2729	2279	76.71	83.51	79.76

Finally, Tables 1.3.5 – 1.3.7 present comparative experimental results for each dataset using the baseline evaluation protocol in terms of BR, BP, and BFM. It should be stressed that we do not include the method FromBaseToPoly due to the fact that it produces only polygons (the method starts from the correct baselines which correspond to the ground truth). In addition to these results, the methods are tested using the training data of the upcoming ICDAR 2017 competition on baseline detection. For this data, a table is presented (Table 1.3.8) which contains the comparative experimental results with respect to each subset of the training dataset as well as with respect to the entire set.

Table 1.3.5: Comparative experimental results using Konzilsprotokolle dataset (baseline evaluation)

Method	# GT lines	# RS lines	BR	BP	BFM
NCSR (1st year)	2555	2532	92.61	91.16	91.88
Louloudis2009	2555	2520	91.56	90.61	91.09

Table 1.3.6: Comparative experimental results using NAF dataset (baseline evaluation)

Method	# GT lines	# RS lines	BR	BP	BFM
NCSR (1st year)	3186	3053	96.05	96.15	96.10
Louloudis2009	3186	3085	96.39	95.00	95.69

Table 1.3.7: Comparative experimental results using BL dataset (baseline evaluation)

Method	# GT lines	# RS lines	BR	BP	BFM
NCSR (1st year)	2971	2889	90.98	88.32	89.63
Louloudis2009	2971	2731	89.08	91.05	90.05

Table 1.3.8: Comparative experimental results using training dataset of baseline detection competition on each subset as well as on the entire dataset (baseline evaluation)

Method	# GT lines	# RS lines	BR	BP	BFM
	ABP_FirstTestCollection				
NCSR (1st year)	961	623	79.12	68.02	73.15
Louloudis2009	961	754	75.94	75.11	75.52
	Bohisto_Bozen_SetP				
NCSR (1st year)	815	651	85.42	76.33	80.62
Louloudis2009	815	787	83.9	86.93	85.39
	EPFL_VTM_FirstTestCollection				
NCSR (1st year)	252	170	74.48	72.59	73.52
Louloudis2009	252	249	62.1	81.93	70.65
	HUB_Berlin_Humboldt				
NCSR (1st year)	693	702	88.96	89.71	89.33
Louloudis2009	693	709	89.13	90.86	89.99
	NAF_FirstTestCollection				
NCSR (1st year)	930	921	92.14	93.84	92.98
Louloudis2009	930	920	91.71	93.38	92.54
	StAM_Marburg_Grimm_SetP				
NCSR (1st year)	856	761	83.87	78.94	81.33
Louloudis2009	856	784	81.35	79.69	80.51
	UCL_Bentham_SetP				
NCSR (1st year)	1024	718	84.51	71.4	77.4
Louloudis2009	1024	937	81.35	83.84	82.58
	unibas_e-Manuscripta				
NCSR (1st year)	848	520	89.21	65.15	75.3
Louloudis2009	848	807	91.9	90.32	91.1
	Entire training dataset				
NCSR (1st year)	6379	5066	85.85	77.49	81.45
Louloudis2009	6379	5947	84.4	85.63	85.01

As the experimental results indicate, the NCSR (1st year) text line segmentation method has a good performance on most datasets. It seems that the performance on the BL dataset is lower mainly due to the erroneous definition of ground truth regions and text lines (see Figure 1.3.1) as well as due to missing characters as a result of the binarization procedure. This is not an issue for the FromBaseToPoly method since this method relies only on the ground truth baselines which are provided as input.

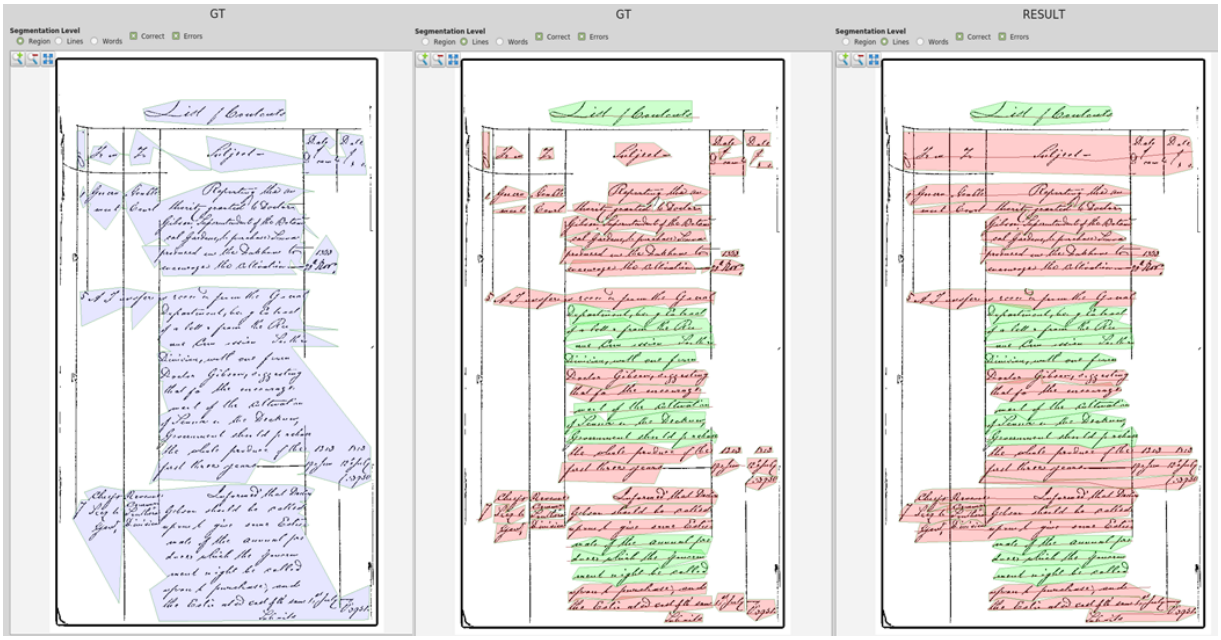


Figure 1.3.1: Indicative result of the NCSR (1st year) method on BL dataset. Erroneous ground truth regions (left image, errors due to merging of regions belonging to different columns) lead to erroneous detections of text lines (red polygons). The middle image corresponds to the ground truth text lines. Finally, the right image corresponds to the automatically produced result.

Concerning the Konzilsprotokolle dataset, the NCSR (1st year) method achieves a polygon F-Measure (PFM) of 90.00% and a baseline F-Measure (BFM) of 91.88%. A representative example using a document of this dataset is shown in Figure 1.3.2 (PFM=92.59%, BFM=93.32%). The two errors are due to the addition of noisy connected components as well as underlines to the final result.

Concerning the results on the training data of the upcoming baseline detection competition, it should be stressed that the modified version (NCSR (1st year)) which tried to solve most of the issues appearing on the “tranScriptorium” datasets (mainly insertions above a text line as well as erroneous splitting of a single text line) seems to fail to generalize on datasets coming from various sources, periods, languages etc. A representative example is shown in Figure 1.3.3.

As the experimental results indicate, the baseline evaluation metrics have significant higher values compared to the polygon evaluation metrics in all comparisons. This can be explained by the fact that the polygon evaluation is a pixel based evaluation which relies on the correct assignment of all pixels to a specific text line. The latter is a stricter rule when compared to the baseline evaluation which relies on the correct positioning of a baseline. An interesting question that should be answered in the next year concerns the degree of correlation of these metrics with the metrics of the handwritten text recognition (HTR) step.

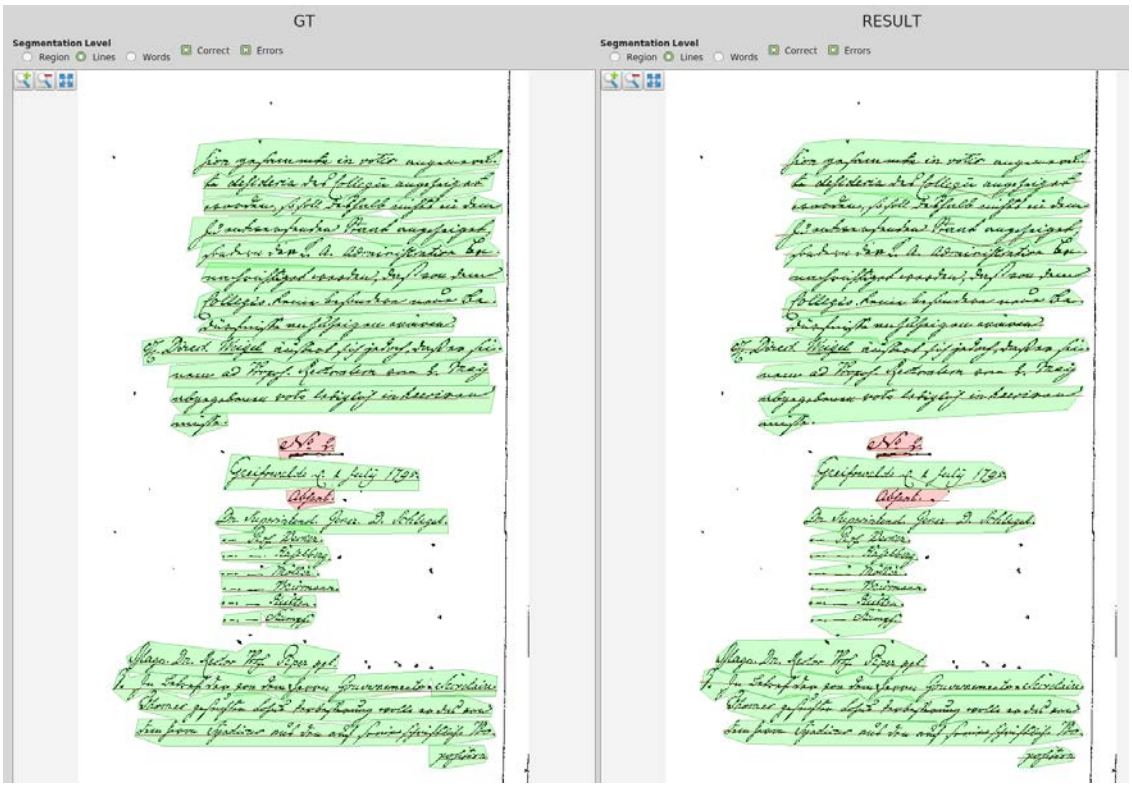
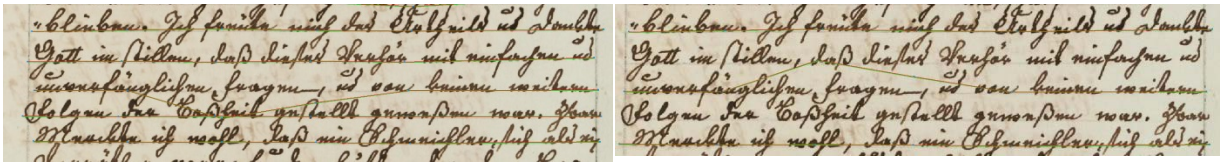


Figure 1.3.2: Indicative result of the NCSR (1st year) method on Konzilsprotokolle dataset. Many of the errors are due to the addition of small parts (noise, underlines) on the result.



(a) **(b)**

Figure 1.3.3: Indicative result of the Louloudis2009 method (a) and NCSR (1st year) method (b) on an image sample of the unibas_e-Manuscripta dataset (subset of dataset for baseline detection). Notice that one baseline is running across two lines in (a). The post-processing step of NCSR 1st year method considers this running as a splitting of a single text line and finally merges this baseline (with the above and below baseline) leading to the erroneous result (b).

In addition, since the baseline representation has the advantage of needing less time for correction and since according to [Romero2015] the baseline representation produces comparable results in terms of HTR accuracy with the polygon representation, we have already started working on a more accurate text line segmentation method which will provide only the baseline representation. We are also motivated to focus on the automatic production of better baselines by the fact that our FromBaseToPoly method can produce very accurate text lines at polygon level as long as it receives accurate baselines.

1.4. CVL Text line Segmentation Method – 1st Year

A brief description of the text line segmentation method developed by CVL can be found in deliverable D6.4 “Basic layout analysis tool”.

1.5. UPVLC Text Line Detection and Classification

1.5.1 Evaluation Protocol

In order to evaluate the quality of the proposed SMSLA approach, we have adopted two types of measures: line error rate (LER) and relative geometric error (RGE).

LER is a qualitative measure that indicates the ratio of regions incorrectly assigned over the total number of regions. The number of incorrectly assigned regions in a page image amounts to the number of label insertions deletions and substitution which have to be done on a vertical layout in order to match the corresponding system hypothesis (h) reference label sequence. It is obtained in the same way as the well-known word error rate (WER) [McCowan2004]; that is, by determining the optimal alignment between the system hypotheses and reference label sequences through dynamic programming. LER is currently the only actual measure that evaluates proper classification of detected text lines into different types.

On the other hand RGE evaluates, in a more quantitative manner, the geometric quality of the detected baseline vertical coordinates with respect to the corresponding reference marks. RGE is computed in two phases. First, for each page image, we find the best alignment between the vertical baseline coordinates yielded by the system and the corresponding reference coordinates for that page. Secondly, we compute the actual RGE as the average (over all lines and pages) of the geometric error in pixels, divided by the average line region height (also in pixels) for the corpus considered. By computing the RGE in this manner me ensure that our measure allows us to compare segmentation quality across corpora with different resolutions and script sizes.

RGE was used at the moment of empirical evaluation for the below described work as at the time it was being developed it was the only baseline based geometric error evaluation method. In future the evaluation method that is being developed inside the READ project will be used.

https://github.com/Transkribus/TranskribusBaseLineMetricTool

1.5.2 Sheet Music Statistical Layout Analysis

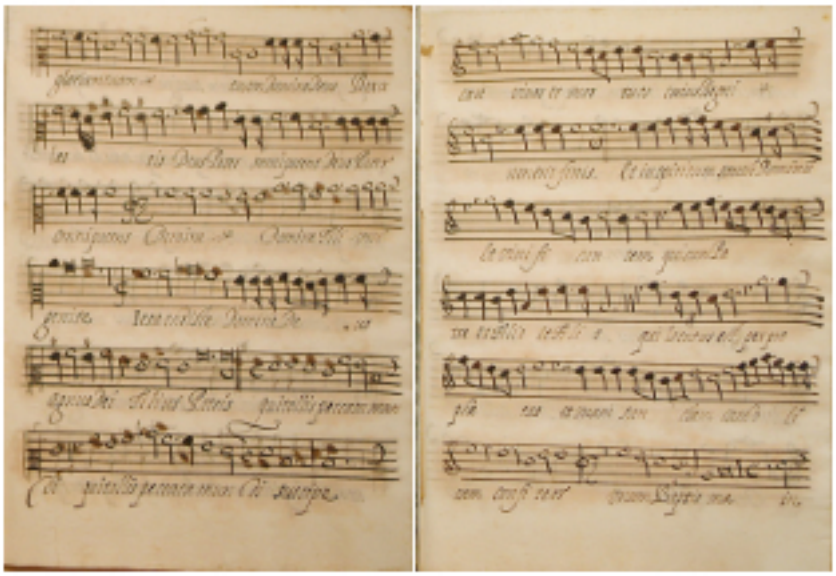
In order to provide access to the contents of ancient music scores to researchers, the transcripts of both the lyrics and the musical notation is required. Before attempting any type of automatic or semi-automatic transcription of sheet music, an adequate layout analysis (LA) is needed. This LA must provide not only the **locations** of the different image regions, but also **adequate region labels** to distinguish between different region types such as staff, lyric, etc.

To this end, we adapted a stochastic framework for LA based on Hidden Markov Models that we had previously introduced for detection and classification of text lines in typical handwritten text images.

The proposed approach takes a scanned music score image as input and, after basic preprocessing, simultaneously performs region detection and region classification in an integrated way.

To assess this statistical LA approach several experiments were carried out on a representative sample of a historical music archive, under different difficulty settings. The results show that our approach is able to tackle these structured documents providing good results not only for region detection but also for classification of the different regions.

The experiments were carried out using a part of the "CAPITÁN", a huge archive of manuscripts of Spanish and Latin American music from the 16th to 18th centuries. These manuscripts were written using the so-called white mensural notation, which in many aspects differ from the modern Western musical notation. Furthermore, this archive was written following the slightly different Hispanic notation of that time, increasing its historical and musicological interest.



The LER and the corresponding RGE are computed for different levels of detail used in the ground-truth labeling. In this work we have studied four levels: detection of foreground regions, Staff and Lyric differentiation, multiple staff sub-classes and multiple lyrics sub-classes. Results are as follows:

Labeling Detail Level	**LER (%)**	**RGE(%)**
Foreground Detection	1.1	3.0
Staff / Lyrics	4.6	3.0
Multiple Lyrics Classes	6.9	3.2
Multiple Staff Classes	28.0	3.9

A more analytical description of this work can be found at [Bosch2016].

1.5.3 Text Line Detection using Clustering

At UPVLC we are currently developing a method for Text Line Detection that uses Extremely Randomized Trees in order to detect the lower contour of written words and combines them by means of a modified version of the DBSCAN clustering algorithm to generate baselines.

This method is already obtaining very good results with both the old polygon F-measure and the new baseline based evaluation method that is being developed inside the READ project with just a couple of pages with marked baselines required. Work to be published during 2017.

2. Word Segmentation

Word segmentation refers to the process of defining the word regions of a text line. Since nowadays most handwriting recognition methods assume text lines as input, the word segmentation process is usually necessary only for segmentation-based query by example (QbE) keyword spotting (KWS) methods. Segmentation of historical handwritten document images still presents significant challenges and it is an open problem. These challenges include the appearance of skew along a single text line, the existence of slant, the non-uniform spacing of words as well as the existence of punctuation marks (Figure 2.1).

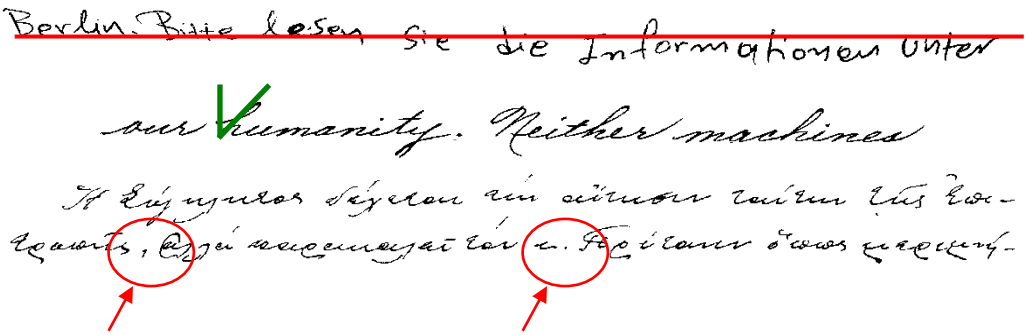
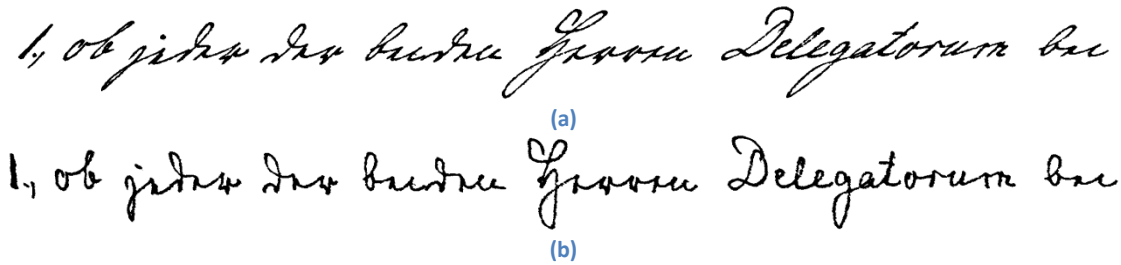


Figure 2.1: Challenges encountered on historical document images for word segmentation.

2.1. NCSR Word Segmentation Method – 1st Year

In the frame of "tranScriptorium" project a novel word segmentation method – NCSR method (1st Year) - was developed which was an extension of the method presented in [Louloudis2009], adapted to historical handwritten documents. The developed method contains two steps. The first step deals with the computation of the Euclidean distances of adjacent components in the text line image and the second step concerns the classification of the previously computed distances as either inter-word gaps or intra-words distances. A more detailed description of the two steps is provided in the sequel.

Distance Computation: In order to calculate the distance of adjacent components in the text line image, a pre-processing procedure is applied. The pre-processing procedure concerns the correction of the dominant slant angle [Vinciarelli2001] of the text line image (Figure 2.1.1). The computation of the distance metric is considered not on the connected components but on the overlapped components (OCs). An OC is defined as a set of connected components whose projection profiles overlap in the vertical direction. The Euclidean distance between two adjacent OCs is defined as the minimum among the Euclidean distances of all pairs of points of the two adjacent OCs.



(a)

(b)

Figure 2.1.1: (a) Original text line image; (b) after slant correction.

Distance Classification: A mixture model clustering is based on the idea that each cluster is mathematically presented by a parametric distribution. We have a two clusters problem (inter-word and intra-word distances) so every cluster is modeled with a Gaussian distribution (Figure 2.1.2). The algorithm that is used to calculate the parameters for the Gaussians is the Expectation Maximization (EM) algorithm. We use this methodology since Gaussian Mixture Modeling is a well-known unsupervised clustering technique with many advantages which include: (i) the mixture model covers the data well, (ii) an estimation of the density for each cluster can be obtained and (iii) a "soft" classification is available. For a detailed description of Gaussian Mixtures, the interested reader is referred to [Marin2005]. For the calculation of the number of parameters and the number of Gaussians the software package "Cluster" was used, which implements an unsupervised algorithm for modeling Gaussian mixtures [https://engineering.purdue.edu/~bouman/software/cluster/].

It should be stressed that the NCSR word segmentation method is developed in C++ following the guidelines of the Transkribus interface and it is available at github:

https://github.com/Transkribus/NCSR_Tools

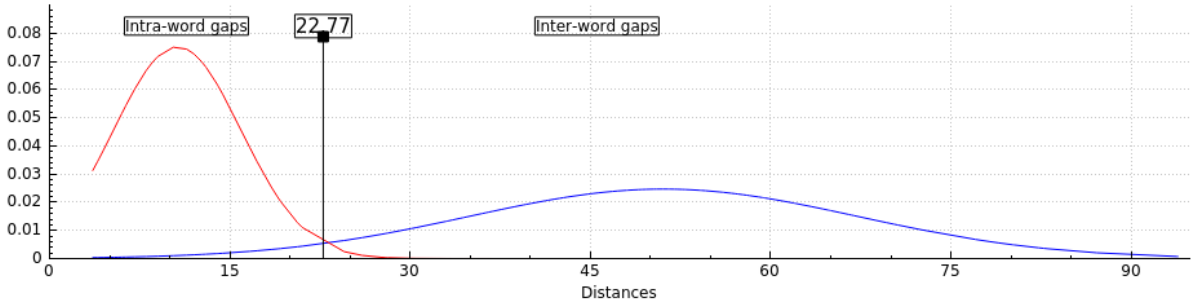


Figure 2.1.2: Example of Gaussian distributions for intra-word and inter-word distances concerning one document image.

2.2. Evaluation

The performance of the NCSR method (1st year) as well as the sequential clustering method [Kim2001] has been tested on three challenging datasets of historical handwritten documents: (i) Konzilsprotokolle (German), (ii) NAF (Finnish) and (iii) BL (English). Table 1.3.1 summarizes the number of documents as well as the number of words for each dataset.

For the evaluation of the word segmentation methods we follow the same protocol which was used in the ICDAR 2013 Handwriting Segmentation Competition [Stamatopoulos2013]. An analytic description of the protocol is provided in the abovementioned section for evaluation of text line segmentation task. The acceptance threshold (T_a) was set to 0.9. Tables 2.2.1 - 2.2.3 present comparative experimental results for each dataset in terms of Precision, Recall, and F-Measure.

Table 2.2.1: Comparative experimental results using Konzilsprotokolle dataset

Method	# GT words	# RS words	#o2o	PP	PR	PFM
NCSR (1st year)	15567	16252	12587	77.45	80.86	79.12
Sequential Clustering	15567	11418	8325	72.91	53.48	61.70

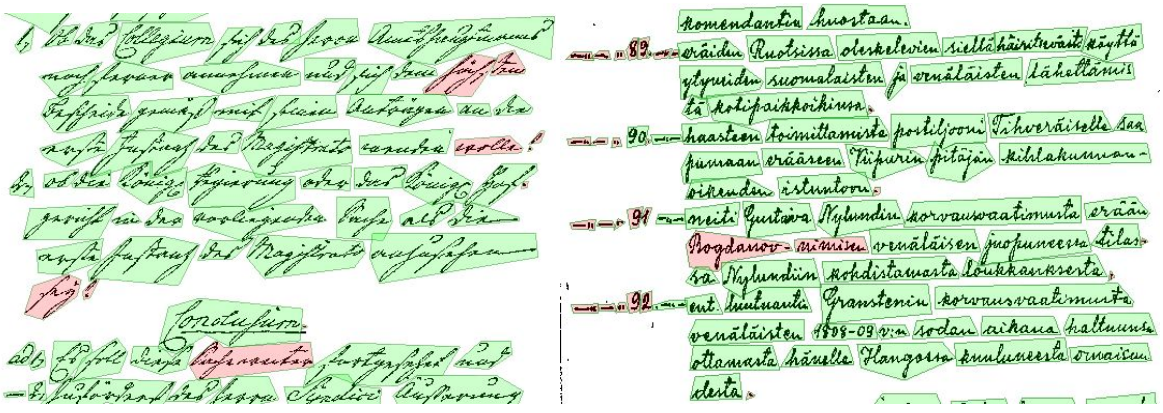
Table 2.2.2: Comparative experimental results using NAF dataset

Method	# GT words	# RS words	#o2o	PP	PR	PFM
NCSR (1st year)	16201	20045	13404	66.87	82.74	73.96
Sequential Clustering	16201	13200	10168	77.03	62.76	69.17

Table 2.2.3: Comparative experimental results using BL dataset

Method	# GT words	# RS words	#o2o	PP	PR	PFM
NCSR (1st year)	15739	16908	11128	65.81	70.70	68.17
Sequential Clustering	15739	11858	8049	67.88	51.14	58.33

As the experimental results indicate, the NCSR (1st year) method outperforms the sequential clustering method on all datasets and it achieves the highest F-Measure on Konzilsprotokolle dataset (79.12%) in which most of the errors have been produced by the punctuation marks (Figure 2.2.1a). Concerning the NAF dataset, the NCSR (1st year) method achieves lower F-Measure (73.96%) since a lot of insignificant errors have been produced due to the presence of ditto marks (Figure 2.2.1b). Finally, the NCSR (1st year) method achieves the lowest performance on the BL dataset (68.17%) since many characters are missing or they are broken due to the binarization procedure (NCSR "tranScriptorium" binarization method). A representative example using a document of the BL dataset is presented in Figure 2.2.2.



(a) (b)

Figure 2.2.1: Indicative results of the NCSR (1st year) method on (a) Konzilsprotokolle dataset in which most of the errors (red polygons) have been produced by the punctuation marks and on (b) NAF dataset in which many insignificant errors (red polygons) have been produced due to the presence of ditto marks.

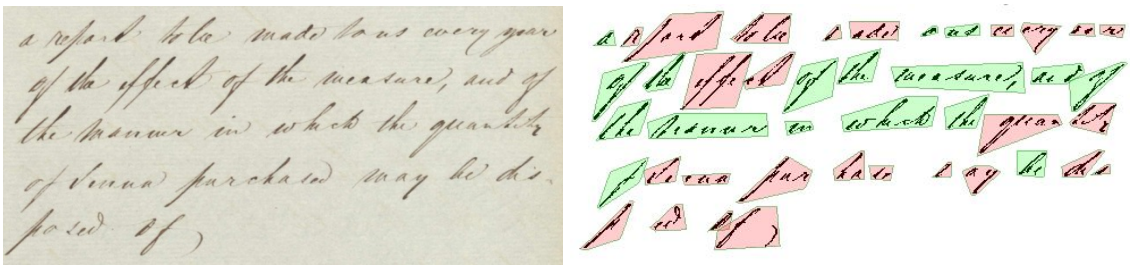
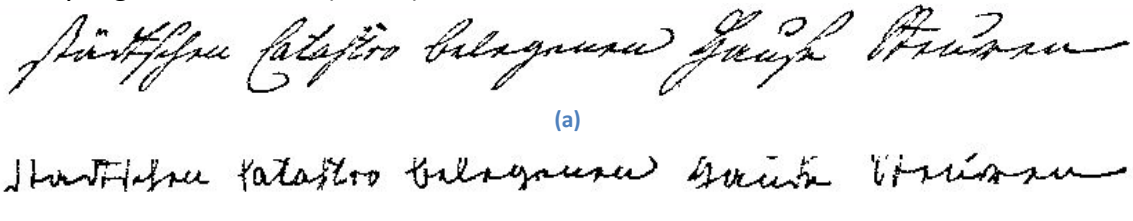


Figure 2.2.2: Indicative result of the NCSR (1st year) method on BL dataset. Many characters are missing or they are broken due to the binarization procedure.

Taking into account the above mentioned observations, we aim to provide a more reliable word segmentation method in order to cope with these challenges. Based on our preliminary experimentation, it seems that main zone detection can be successfully applied in order to exclude the ascenders/descenders as well as the punctuation marks from the distance computation step (see Figure 2.2.3). Baseline information provided by the text line segmentation procedure can be used in order to define the main zone. Moreover, concerning the distance classification step, we have performed experiments using different techniques, instead of a Gaussian distribution, such as the Student's-t distribution. The main advantage of the Student's-t distribution concerns its robustness to the existence of outliers. Finally, since the word segmentation process is usually necessary only for segmentation-based query by example keyword spotting methods, we will investigate the scenario of providing multiple hypothesis segmentation results in order to increase the number of correctly segmented words (Recall).



(a)

(b)

Figure 2.2.3: Example of main zone detection in order to exclude the ascenders/descenders as well as the punctuation marks; (a) original text line; (b) main zone information.

3. References

[Bosch2016] V. Bosch, J. C.-Zaragoza, A. Toselli and E. Vidal, "Sheet Music Statistical Layout Analysis", 15th International Conference on Frontiers in Handwriting Recognition (ICFHR'16), pp. 313-318, 2016.

[Gatos2014] B. Gatos, G. Louloudis and N. Stamatopoulos "Segmentation of Historical Handwritten Documents into Text Zones and Text Lines", 14th International Conference on Frontiers in Handwriting Recognition (ICFHR'14), pp. 464-469, 2014.

[Gatos2015] http://transcriptorium.eu/pdfs/deliverables/tranScriptorium-D3.2.2-31August2015.pdf

[Gruning2016] https://github.com/Transkribus/TranskribusBaseLineMetricTool

[Huang2008] C. Huang and S. Srihari, "Word segmentation of off-line handwritten documents", Proc. Annual Symposium on Document Recognition and Retrieval (DRR) XV, IST/SPIE, 2008.

[Kim2001] S.H. Kim, S. Jeong, G.S. Lee and C.Y. Suen, "Word segmentation in handwritten Korean text lines based on gap clustering techniques", 6th International Conference on Document Analysis and Recognition (ICDAR'01), pp. 189-193, 2001.

[Louloudis2009] G. Louloudis, B. Gatos, I. Pratikakis, C. Halatsis, "Text line and word segmentation of handwritten documents", Pattern Recognition, vol. 42, no 12, pp. 3169-3183, 2009.

[Marin2005] J.M. Marin, K. Mengersen and C.P. Robert, "Bayesian Modelling and Inference on Mixtures of Distributions", Handbook of Statistics, vol. 25, Elsevier-Sciences, 2005.

[McCowan2004] I. A. McCowan, D. Moore, J. Dines, D. Gatica-Perez, M. Flynn, P. Wellner, and H. Bourlard, "On the use of information retrieval measures for speech recognition evaluation," IDIAP, Martigny, Switzerland, Idiap-RR Idiap-RR-73-2004, 2004.

[Retsinas2016] G. Retsinas, G. Louloudis, N. Stamatopoulos and B. Gatos, "Efficient Document Image Segmentation Representation by Approximating Minimum-Link Polygons", 12th Workshop on Document Analysis Systems (DAS'16), pp. 293-298, 2016.

[Romero2015] V. Romero, J.A. Sanchez, V. Bosch, K. Depuydt, and J. de Does, "Influence of text line segmentation in handwritten text recognition", 13th International Conference on Document Analysis and Recognition, pp. 536-540, 2015.

[Stamatopoulos2013] N. Stamatopoulos, G. Louloudis, B. Gatos, U. Pal and A. Alaei, "ICDAR2013 Handwriting Segmentation Contest", International Conference on Document Analysis and Recognition (ICDAR'13), pp. 1402-1406, 2013.

[Vinciarelli2001] A. Vinciarelli and J. Juergen, "A new normalization technique for cursive handwritten words", Pattern Recognition Letters, vol. 22, num. 9, pp. 1043-1050, 2001.