UNIVERSITAT POLITÈCNICA DE VALÈNCIA
INSTITUTO TECNOLÓGICO DE INFORMÁTICA

PATTERN RECOGNITION AND
HUMAN LANGUAGE TECHNOLOGY GROUP

IARFID Master Thesis

Interactive Layout Analysis

Author: Lorenzo Quirós Díaz

Advisor: Carlos D. Martínez Hinarejos
Co-advisor: Alejandro Héctor Toselli
Co-advisor: Enrique Vidal Ruiz

September 12, 2016

Abstract

The amounts of ancient documents transcribed by means of HTR technology have been rising dramatically over the last years. Consequently, the develop and enhancement of HTR methods and algorithms have been rising as well. However, Layout Analysis remains a bottleneck in the development and generalization of HTR technology. In this work a new Interactive-Probabilistic method is presented. This new method incorporates the user feedback in the Layout Analysis process, in order to provide, not just a very accurate layout but an interactive framework in which user feedback is used to help the user to fix any error.

Resumen

La cantidad de documentos antiguos transcritos por medio de la tecnología HTR se han incrementado dramáticamente en los últimos años. En consecuencia, el desarrollo y la mejora de los métodos y algoritmos de HTR han estado aumentando también. Sin embargo, *Layout Analysis* sigue siendo un cuello de botella en el desarrollo y la generalización de la tecnología de HTR. En este trabajo se presenta un nuevo método Interactivo-Probabilístico. Este nuevo método incorpora la retroalimentación del usuario en el proceso de *Layout Analysis*, con el fin de proporcionar, no sólo un diseño muy preciso, sino un marco interactivo en el que se utiliza la retroalimentación de los usuarios para ayudar al usuario a corregir cualquier error.

Resum

La quantitat de documents antics transcrits per mitjà de la tecnologia HTR s'han incrementat dramàticament en els últims anys. En conseqüència, el desenvolupament i la millora dels mètodes i algoritmes de HTR han estat augmentant també. No obstant això, *Layout Analysis* segueix sent un coll d'ampolla en el desenvolupament i la generalització de la tecnologia de HTR. En aquest treball es presenta un nou mètode Interactiu-Probabilístic. Aquest nou mètode incorpora la retroalimentació de l'usuari en el procés de *Layout Analysis*, per tal de proporcionar, no només un disseny molt precís, sinó un marc interactiu en què s'utilitza la retroalimentació dels usuaris per ajudar l'usuari a corregir qualsevol error.

Acknowledgements

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Chapter Outline

1.1 Introduction

Handwritten Text Recognition can be defined as the problem of finding the most likely work sequence, for a given handwritten sequence image [16]. Under this definition the presence of a input image is needed, but this image should contain only a handwritten sequence, this is a single "line" of handwritten text should be in any image at the time. Now, since main goal of HTR systems is to translate not just a single line, but the complete page or even the complete book, a previous system is needed in order to extract those lines from the whole page and, in upper level, to extract the different zones of the page (paragraph, marginal notes, illustrations, page number, etc.). Notice, this is a very important stage on the HTR process; for example, if we develop a system to recognize text in a image, but we provide an illustration without text, the results of that system are expected to be erroneous, or in the best case system is capable to ignore that kind of inputs, but we expend time and computing resources on it. In addition, knowledge of the type of zone is very useful to provide some context to the text in the zone, and because each zone can be semantically different, each zone can be processed differently.

Document Layout analysis (DLA) is the process of identifying and categorizing the regions of interest in a image of document. Commonly this process is divided in two sub problems [6]: Detection and labeling of the different zones in the image (body, illustrations, marginalia) is called *geometric layout analysis*, and the classification of those zones into their logical role (title, caption, footnote, etc.) is called the *logical layout analysis*. Several methods have been developed for DLA [1–3, 5, 9, 15, 18], most of them based on Computer Vision techniques, for instance, binarization [11, 13], skew correction [10, 12, 14], Connected Components labeling [4, 8], among others. All of those methods are designed as user-free system, as a consequence, any error on the system result must be fixed from scratch by the user.

On this work an Interactive-Probabilistic approach for image layout analysis is proposed, on the aim to provide, not just a very accurate layout but an interactive framework in which user feedback is used to fix any error. Conditional Random Fields model have been combined by prior-probability model and Interactive Pattern Recognition [17] to build the new method.

This document is organized as follows: In the first chapter we introduce the topic under study, and cover the motivation and our approach. On Chapter 2 we cover technical fundaments of our approach along with evaluation measures.

Our approach is explained in detail in Chapter 3. Then, experiments and results that help us to evaluate proposed method are presented and discussed on Chapter 4. Finally, in Chapter 5 we present the conclusions extracted from our investigation, and future work to be covered.

1.2 Motivation

Several manuscripts have being digitized in order to preserve the valued information from the day-by-day deterioration of the physical document. The amount and importance of the information contained on those manuscripts motivates the development of several techniques and tools to explore, analyze and read them in a more comprehensive manner. This is from image quality enhancement to automatic Handwritten Text Recognition (HTR).

One of the most crucial steps in HTR is Document Layout Analysis, this is the process by which zones of interest in a page image are detected and categorized. This process it is commonly performed by hand, and some times assisted by semi-automatic methods [7, 15, 19] (so called semi-automatic because user always needs to review the results).

With the premise that the user always needs to review[a] the results, and the importance of the DLA in HTR systems, is highly desirable to study interactive methods to improve the aforementioned systems.

1.3 Related Work

Several methods have been developed on the last few years to extract the correct Layout from digitized pages, most of them [1–3, 5, 9, 15, 18] follows a similar set of steps:

(i) Image binarization.

(ii) Skew correction.

(iii) Connected Components and/or White spaces analysis.

(iv) Some heuristic to link/merge blocks found by previous step.

(v) Clean the result (filter out small blocks, remove unsuitable blocks, etc.).

[a]We hope some day an completely automatic system could arrive, but at the time this paper is written state of the art methods are far away of that goal.

Furthermore, all of the methods studied a binarization step is performed, which is a very hard problem by itself, this result in an increase in the difficulty of the original problem and a direct dependence from the binarization method. Also, heuristics performed limit the generalizability of the methods, since several parameters needs to be adjusted by the user based on the characteristics of the input images (page degradation, number of columns, illustrations, quality of the scan, etc.).

1.4 Overview of the Proposal Approach

As mentioned in above sections, methods proposed until now sill have errors, this is the user needs to go trough all the images in order to review if the layout proposed by the classification method is correct or not, then if there is any error user must fix it. On this work we proposed to use a probabilistic methodology to provide not just a good layout but to help the user to easily fix any error provided by the classifier. Under this context, we use Interactive Pattern Recognition (IPR) framework proposed by Toselli et al. [17] to solve the issue presented on Layout Analysis.

IPR framework forces, from its definition, to use a probabilistic model, in order to take advantage of the information inside the model and the information provided by the user iteratively. This is, we do not need just a classification of the zones of the page, ie. a single hypothesis is provided by the classifier; instead, we need the probability of each possible hypothesis, were each hypothesis represent a possible layout under the page restrictions. Of course, this normally meas a huge search space, even on a simple small image. In order to guide search algorithms to the best hypothesis, prior-probability distribution over the corners of the zones should be learned from training data.

We use Conditional Random Fields to model the conditional distribution of the image pixels to each layout zone ($p(h_k|x_{ij})$) and provide the main probability distribution to IPR framework, also multi-variable Gaussian mixture models are used to learn the prior-probability distribution of the corners of each zone in the layout.

Finally, brute force approach is taken to search best hypothesis under probabilistic model and the user feedback.

1.5 Context of Applications and Assumptions

Many enhancements and and test cases should be done to the proposed system before reach an stable version of it. For this reason project will be divided on some stages, on this work we will present the only the first one, which framed by the following constrains for experimentation:

- Experiments will be conduced over a small set of images.
- Experiments will be limited to search for only the main paragraph in the pages.
- Page images are assumed to have only a main paragraph, marginalia, titles, catch-words, and other types of zones, but not illustrations or figures.

These constrains are adequate to present the method without loss of generality, because all the theory behind is not limited by the constrains, just the experimentation.

1.6 Expected Outcomes/Results

The expected results are:

- A formal definition of the new probabilistic method, based on CRFs, GMMs and IPR framework.
- A system that obtain competitive results on the task.
- Review the impact of the new approach in the final user effort.
- A demo to present new method features.
- Obtain a solid base for future project steps.

Bibliography

[1] Antonacopoulos, A., Clausner, C., Papadopoulos, C., and Pletschacher, S. (2011). Historical document layout analysis competition. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, pages 1516–1520.

[2] Antonacopoulos, A., Clausner, C., Papadopoulos, C., and Pletschacher, S. (2013). ICDAR 2013 competition on historical book recognition (HBR 2013). *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, pages 1459–1463.

[3] Antonacopoulos, A., Clausner, C., Papadopoulos, C., and Pletschacher, S. (2015). 2015_ICDAR2015 Competition on Recognition of Documents with Complex Layouts -RDCL2015. In *Document Analysis and Recognition (ICDAR), 2015 13th International Conference on*, pages 1151–1155.

[4] Bieniek, A. and Moga, A. (2000). An efficient watershed algorithm based on connected components. *Pattern Recognition*, 33(6):907–916.

[5] Bukhari, S. S., Breuel, T. M., Asi, A., and El-Sana, J. (2012). Layout analysis for Arabic historical document images using machine learning. *Proceedings - International Workshop on Frontiers in Handwriting Recognition, IWFHR*, pages 639–644.

[6] Cattoni, R., Coianiz, T., Messelodi, S., and Modena, C. M. (1998). Geometric layout analysis techniques for document image understanding: a review. Technical report.

[7] Gatos, B., Stamatopoulos, N., and Louloudis, G. (2011). ICDAR2009 handwriting segmentation contest. *International Journal on Document Analysis and Recognition*, 14(1):25–33.

[8] Grana, C., Borghesani, D., and Cucchiara, R. (2010). Optimized block-based connected components labeling with decision trees. *IEEE Transactions on Image Processing*, 19(6):1596–1609.

[9] Lazzara, G., Geraud, T., and Levillain, R. (2014). Planting, growing, and pruning trees: Connected filters applied to document image analysis. *Proceedings - 11th IAPR International Workshop on Document Analysis Systems, DAS 2014*, pages 36–40.

[10] Liu, H., Wu, Q., Zha, H., and Liu, X. (2008). Skew detection for complex document images using robust borderlines in both text and non-text regions. *Pattern Recognition Letters*, 29(13):1893–1900.

[11] Niblack, W. (1985). *An Introduction to Digital Image Processing*. Strandberg Publishing Company.

[12] Papandreou, A., Gatos, B., Louloudis, G., and Stamatopoulos, N. (2013). ICDAR 2013 document image skew estimation contest (DISEC 2013). *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, pages 1444–1448.

[13] Sauvola, J. and Pietikäinen, M. (2000). Adaptive document image binarization. *Pattern Recognition*, 33(2):225–236.

[14] Singh, C., Bhatia, N., and Kaur, A. (2008). Hough transform based fast skew detection and accurate skew correction methods. *Pattern Recognition*, 41(12):3528–3546.

[15] Smith, R. (2009). Hybrid Page Layout Analysis via Tab-Stop Detection. In *Proceedings of the 2009 10th International Conference on Document Analysis and Recognition*, pages 241–245. IEEE Computer Society.

[16] Toselli, A. H., S, J. A., Vidal, E., Polit, U., and Val, D. (2011a). Handwritten Text Recognition for Historical Documents. (September):90–96.

[17] Toselli, A. H., Vidal, E., and Casacuberta, F. (2011b). *Multimodal Interactive Pattern Recognition and Applications*. Springer.

[18] Vil'kin, A. M., Safonov, I. V., and Egorova, M. A. (2013). Algorithm for Segmentation of Documents Based on Texture Features. *Pattern Recognit. Image Anal.*, 23(March, 2013):153–159.

[19] Wang, S. Y., Baird, H. S., and An, C. (2009). Document content extraction using automatically discovered features. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, pages 1076–1080.

CHAPTER 2

FUNDAMENTS

Chapter Outline

2.1 Fundaments [WIP]

2.2 Principal Components Analysis [WIP]

Algorithm 1: PCA basic algorithm.

Data: $n, d, k, \{1_1, x_2, \cdots, x_n\}$
Result: $\{W, \overline{x}\}$

2.3 Conditional Random Fields [WIP]

Let's define a general prediction problem where is required to model the existing dependencies between the set of random variables $X \cup Y$, considering X as the set of real-valued feature variables and Y the set of class variables to predict. This is a common problem on pattern recognition and machine learning and several methods have been developed towards

Conditional Random Fields (CRF) is a popular probabilistic method for structured prediction, based on graphical modeling [8].

2.4 Interactive Pattern Recognition [WIP]

$$\hat{h} = \arg\max_{h \in \mathcal{H}} P(h|x, h', d) = \arg\max_{h \in \mathcal{H}} P(x|h', d, h) P(h, h', d) \tag{2.1}$$

2.5 Gaussian Mixture Models [WIP]

2.6 Evaluation Measures

Evaluation methodologies for quantitative and qualitative results in any scientific or engineering study are fundamental in order to measure study characteristics and provide a defined and structured way to be compared with others. Layout analysis lacks on a single well defined and widespread evaluation methodology, instead several methods are available on the literature most of them developed to text line detection. Pixel-wise evaluation (sometimes called Overlap-based) have been used by [1, 2, 5], but pixel-wise evaluation is highly

dependent of ground-through, for that reason other methods have been developed to minimize the effect of the ground-trough on the evaluation metric [3, 7] or just to measure performance based on the zones instead the pixels [6].
On this work, general pixel-wise precision/recall methodology is used to evaluate CRFs performance, since no zone is detected but the sites that belonging to specific zone. For methods were zone is detected MatchScore [6] and Goal-Oriented Success Rate (GoSR) [7] will be used.

2.6.1 Pixel-wise performance

Pixel-wise performance will be computed by means of the well known performance (P) Eq. 2.2, recall (R) Eq. 2.3 and F1-score (F1) Eq. 2.4.

$$P = \frac{TP}{TP + FP} \tag{2.2}$$

$$R = \frac{TP}{TP + FN} \tag{2.3}$$

where TP, FP and FN stands for "True Positives", "False Positives" and "False Negatives" respectively.

$$F1 = \frac{2 \cdot P \cdot R}{P + R} \tag{2.4}$$

2.6.2 MatchScore

MatchScore were used on ICDAR (2005 to 2009) competitions [4]. Method is intended, first, to select if a zone pair (h_i^*, h_k) is a one-to-one match Eq. 2.5, ie. the ground-trough zone (h_i^*) match to result zone (h_k) based on some threshold T_a, then if the zones have a one-to-one match performance metric is computed as shown in Eq. 2.6.

$$MatchScore(i,k) = \frac{T(h_i^* \cap h_k)}{T(h_i^* \cup h_k)} \tag{2.5}$$

where $T(\cdot)$ a function which counts the pixels inside a zone.

$$FM_{i,k} = \frac{2T(h_i^* \cap h_k)}{T(h_i^*) + T(h_k)} \tag{2.6}$$

Finally global performance is extracted by calculating the average values of performance metric for all one-to-one zone pair.

2.6.3 Goal-Oriented Success Rate

Goal-Oriented performance aims to evaluate how much of the information contained in the ground-trough is also contained in the system result, due that, only foreground pixels are taken into account. Use of foreground pixels forces us to have a pixel level ground-trough, most of the times in the form of a binarized version of the image. Also methodology to define if some zone have one-to-one match is changed from MatchScore to Eq. 2.7.

$$I_{ik} = \begin{cases} h_i^* \cap h_k & \text{if } h_i^* \cap h_k \neq \emptyset \text{ and } \frac{t(I_{ik})}{T(h_k)} > T_a \\ \emptyset & otherwise \end{cases} \tag{2.7}$$

Then, success rate (SR) is defined as follows:

$$SR = \frac{\sum_{i=1}^{|h^*|} \sum_{k=1}^{K} w_{ik} \times T(I_{ik})}{\sum_{i=1}^{|h^*|} T(h_i^*)} \tag{2.8}$$

where w_{ik} corresponds to a weight for each intersection region I_{ik} ranging the interval $[0, ..., 1]$, and depends of the following conditions[a]:

(i) the ground-truth region h_i^* has been detected correctly

(ii) the ground-truth region h_i^* has been split

(iii) the result region h_k has been overlapped by two or more ground-truth regions (merge)

(iv) non-text elements have been included in the result region h_k

(v) if more than one of the previous conditions is satisfied, the weight with the smaller value is selected

Notice that in our case since only one zone is under scrutiny (the main paragraph), any other zone in the layout is considered as non-text element, therefore, w_{ik} belongs to condition (iv), hence, to Eq. 2.9.

$$w_{ik} = \frac{T(I_{ik})}{T(h_k)} \tag{2.9}$$

Bibliography

[1] Baechler, M. and Ingold, R. (2011). Multi Resolution Layout Analysis of Medieval Manuscripts Using Dynamic MLP.

[a] A full list of equations for each condition can be found on [7]

[2] Chaudhury, S., Jindal, M., and Roy, S. D. (2009). Model-Guided Segmentation and Layout Labelling of Document Images using a Hierarchical Conditional Random Field. In *International Conference on Pattern Recognition and Machine Intelligence*, pages 375–380. Springer Berlin Heidelberg.

[3] Clausner, C., Pletschacher, S., and Antonacopoulos, A. (2011). Scenario driven in-depth performance evaluation of document layout analysis methods. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, pages 1404–1408.

[4] Gatos, B., Stamatopoulos, N., and Louloudis, G. (2011). ICDAR2009 handwriting segmentation contest. *International Journal on Document Analysis and Recognition*, 14(1):25–33.

[5] Nicolas, S., Paquet, T., and Heutte, L. (2006). A Markovian approach for handwritten document segmentation. *Proceedings - International Conference on Pattern Recognition*, 3:292–295.

[6] Phillips, I. T., Chhabra, A. K., and Member, S. (1999). of Graphics Recognition Systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(9):849–870.

[7] Stamatopoulos, N., Louloudis, G., and Gatos, B. (2015). Goal-Oriented Performance Evaluation Methodology for Page Segmentation Techniques. In *13th International Confrence on Document Analysis and Recognition - ICDAR'15*, pages 281–285.

[8] Sutton, C. and McCallum, A. (2010). An Introduction to Conditional Random Fields. *Machine Learning*, 4(4):267–373.

CHAPTER 3

INTERACTIVE LAYOUT ANALYSIS

Chapter Outline

3.1 Interactive Layout Analysis

Document Layout analysis is the process of identifying and categorizing the regions of interest in a image of document. Commonly this process is divided in two sub problems [1]: Detection and labeling of the different zones in the image (body, illustrations, marginalia) is called geometric layout analysis, and the classification of those zones into their logical role (title, caption, footnote, etc.) is called the logical layout analysis.
As an input we have an image $\mathcal{X} = \{x_{1,1}, x_{1,2}, .., x_{n,m}\}$ which is associated with rectangular grid $\mathbf{G}$ of size $n \times m$. Each image site s is associated to a cell on the grid defined by its coordinates over $\mathbf{G}$ and denoted $\mathbf{G}_{ij}, \quad 1 \leq i \leq n \quad 1 \leq j \leq m$. The site set is denoted $S = \{s_1, s_2, ..., s_D\} \quad 1 \leq D \leq n \times m$.

Each image is associated to some K-zones Layout defined as $L = \{l_1, l_2, ..., l_K\}$. Then each site s_d belongs to a single layout zone, this is $s_d \Rightarrow l_k; \quad l_k \in L \quad s_d \in S$.
In our case each layout zone is a rectangle defined by its coordinates over G as $l_k = (\boldsymbol{u}_k, \boldsymbol{b}_k) \quad 1 \leq k \leq K$. Figure 3.1 shows an example of this kind of layout zones.

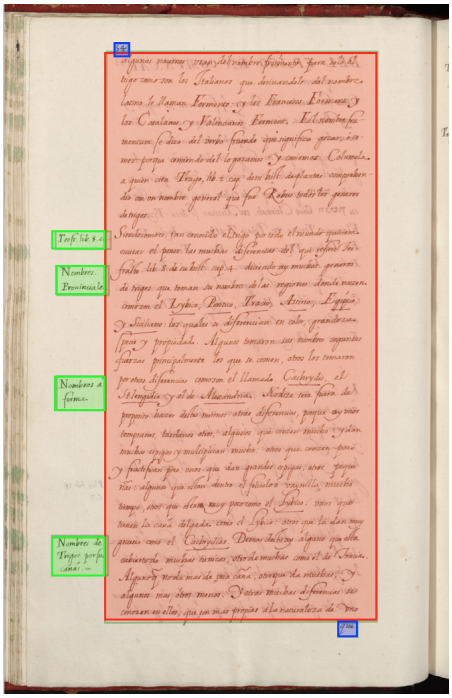


Figure 3.1: Layout zones example. Red=Paragraph, Green=marginalia, Blue=catchword

3.1.0.1 Layout Analysis Problem

Let's define a *structured hypotheses space* [3] $\mathcal{H} = \{h_1, h_2, ..., h_T\}$[a] over the site set S, where $h_t \subseteq L, \quad 1 \leq t \leq T$. We want the hypothesis $\hat{h}$ which provides the best layout for the site set. Under *minimal error criterion*, a best hypothesis is shown to be one which maximizes the posterior probability [2] $P(h|S)$.

$$\hat{h} = \arg\max_{h \in \mathcal{H}} P(h|S) \tag{3.1}$$

However, in many cases is difficult to directly estimate $P(h|S)$ and is better to apply the Bayes rule [3]:

$$\hat{h} = \arg\max_{h \in \mathcal{H}} \frac{P(S|h)P(h)}{P(S)} = P(S|h)P(h) \tag{3.2}$$

where the term $P(S)$ has been dropped since it does not depend on the maximization variable, h. $P(S|h)$ is the probability of an site set S, given the layout hypothesis h and $P(h)$ is their prior probability.

Prior probability $P(h)$ can be modeled by Gaussian Mixture Model ($\mathcal{M}_g$), this is a Gaussian mixture for each corner of each zone layout, where each corner is independent of the others but constrained, by computational reasons, to $\boldsymbol{u}_k > \boldsymbol{b}_k \forall k \in h$.

$$P(h) \approx P_{\mathcal{M}_g}(h) = \prod_{k=1}^{K} P_{\mathcal{M}_g}(\boldsymbol{u}_k) P_{\mathcal{M}_g}(\boldsymbol{b}_k) \tag{3.3}$$

The likelihood, $P(S|h)$, can be approached thought a simple naive Bayes decomposition,under spatial independence assumption[b]; as follows:

$$P(S|h) = \prod_{d=1}^{D} P(s_d|h) \tag{3.4}$$

[a] Based on this definition $\mathcal{H}$ is finite but huge, in the worst case each site s belongs to different layout zone, then $|\mathcal{H}|$ is defined by the size of the set sites and the number of different zones as:

$$|\mathcal{H}| = \binom{|L|+D-1}{D} = \frac{(|L|+D-1)!}{D!(|L|-1)!}$$

[b] Spatial Independence: each site is independent from the others in the set

where each $P(s_d|h)$ can be modeled, for instance, by K-NN, CRF's, RBMs+linear regression function, NN, etc

Formally, Eq. 3.2 can be re-written as:

$$\hat{h} \approx \arg\max_{h \in \mathcal{H}} \prod_{k=1}^{K} P_{\mathcal{M}_g}(\boldsymbol{u}_k) P_{\mathcal{M}_g}(\boldsymbol{b}_k) \prod_{d=1}^{D_k} P(s_d|(\boldsymbol{u}_k, \boldsymbol{b}_k)) \tag{3.5}$$

where D_k is the sub-set of sites inside the zone layout k. In order to prevent precision issues on our calculations we apply $\log$ in both sides of Eq. 3.5.

$$\log \hat{h} \approx \arg\max_{h \in \mathcal{H}} \sum_{k=1}^{K} \left(\log P_{\mathcal{M}_g}(\boldsymbol{u}_k) + \log P_{\mathcal{M}_g}(\boldsymbol{b}_k) + \sum_{d=1}^{D_k} \log P(s_d|(\boldsymbol{u}_k, \boldsymbol{b}_k)) \right) \tag{3.6}$$

3.1.1 Interactive Framework

Interactive framework aims to improve classical pattern recognition paradigm results by taking into account user feedback. As Layout Analysis problem is defined in Eq. 3.6 system will provide the best hypothesis it can, but if there is any error, user commonly will need to remove the prediction and define layout from scratch. This kind of issues could be minimized by the insertion of the feedback (f) and the feedback history (h') on the model [3]. From Eq. 2.1 into Eq. 3.6 a interactive version of the system is defined as:

$$\log \hat{h} \approx \arg\max_{h \in \mathcal{H}} \sum_{k=1}^{K} \left(\log P_{\mathcal{M}_g}(\boldsymbol{u}_k|h', f) + \log P_{\mathcal{M}_g}(\boldsymbol{b}_k|h', f) + \sum_{d=1}^{D_k} \log P(s_d|(\boldsymbol{u}_k, \boldsymbol{b}_k), h', f) \right) \tag{3.7}$$

as a result, the new system must be designed to compute Eq. 3.6 as a first hypothesis, then Eq. 3.7 will be computed on each iteration until hypothesis presented satisfies the user.

3.2 System Architecture [WIP]

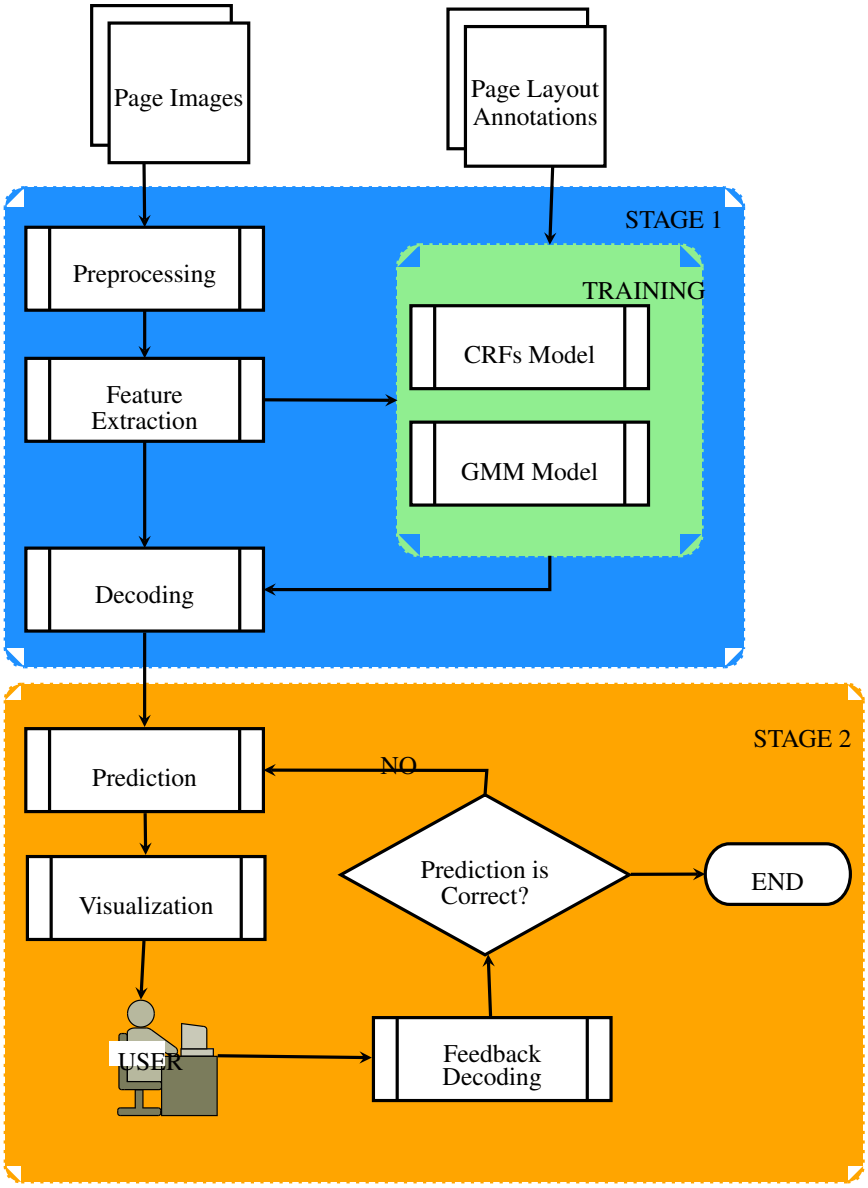


Figure 3.2: System Architecture Diagram

3.3 Pre-processing

In order to keep this stage of the project as simple as possible, and under the assumption that pre-processing images will provide an enhancement over system performance, images are only transformed to intensity images and resolution reduced by feature extraction constrains. This pre-processing steps are focused

on reduce computational cost instead of improve system performance. Other well know techniques, such as, noise reduction, binarization, skew correction, would be studied on next project stages.

3.4 Feature Extraction [WIP]

Feature selection is a very important step on CRF's systems, but also have a direct affect on the CRF model size and training time. Taking those in mind, and in order to keep the system as simple es possible on this stage, the following attributes are selected:

1. PCA-reduced matrix of the site, and their $(w \times w)$ neighborhood, as defined in Eq. 3.8.

$$\mathrm{Q}(s_d) = PCA\left(PCA\left(\mathbf{G}[i-\frac{w}{2}:i+\frac{w}{2}, j-\frac{w}{2}:j+\frac{w}{2}], 3\right)^T, 3\right)\Big|_{i=\lfloor\frac{d}{m}\rfloor, j=d-\lfloor\frac{d}{m}\rfloor} \tag{3.8}$$

   Notice that we are not working over all sites and their respective neighbors, but over a single site and its respective neighbors each time. Hence, we are searching for a reduced representation of the specific site, not a reduced representation of all the sites.

2. Absolute position of the site in the image. This is for each site s_d the row ($\lfloor\frac{d}{m}\rfloor$) and the column ($d - \lfloor\frac{d}{m}\rfloor$) in $\mathbf{G}$.

Then features should be a function of those attributes and the format of how that feature is represented depends of the tool used to learn the CRF model. See Section 3.5.

3.5 CRF's Learning [WIP]

$$\begin{aligned} f_l(s_d, s_{d-1}, s_{d+1}, s_{d-m}, s_{d+m}) = \mathrm{Q}(s_d) \wedge \mathrm{Q}(s_{d-1}) \wedge \mathrm{Q}(s_{d+1}) \wedge \mathrm{Q}(s_{d-m}) \\ \wedge \mathrm{Q}(s_{d+m}) \wedge \mathrm{Q}(s_{d-1})|\mathrm{Q}(s_d) \wedge \mathrm{Q}(s_d)|\mathrm{Q}(s_{d+1}) \\ \wedge \mathrm{Q}(s_{d-m})|\mathrm{Q}(s_d) \wedge \mathrm{Q}(s_d)|\mathrm{Q}(s_{d+m}) \\ \wedge \lfloor\frac{d}{m}\rfloor \wedge d - \lfloor\frac{d}{m}\rfloor \end{aligned} \tag{3.9}$$

3.6 GMM Learning [WIP]

Prior-probability model of where each corner of the zones is in the image, can be estimated by a multi-variable GMM. This is, each corner in the ground-through (h^*) is used to estimate the GMM

Many tools have been developed to estimate GMMs from data, using the EM algorithm. For Python a library called scikit-learn is particularly well known library which implements GMM estimation among many other algorithms, it is very easy to use and efficient, hence is used in this word to estimate GMMs of each corner of each zone layout. On Figure 3.3 log-probability estimated by GMMs of main paragraph corners is drawn for all possible $\boldsymbol{u_k}$ and $\boldsymbol{b}_k$ (k equals paragraph in this case) in an image.

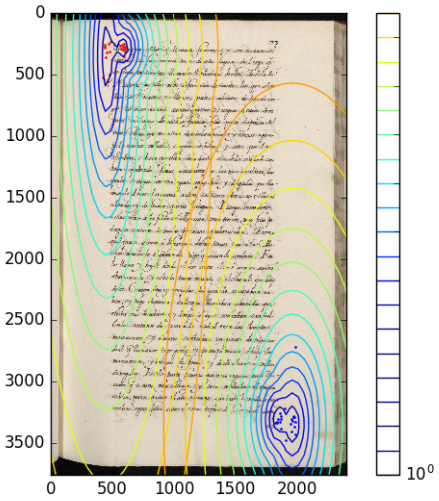


Figure 3.3: Gaussian Mixture Models over main paragraph corners.

3.7 User Interaction

As defined on previous sections, user will review all the pages and provide some feedback to the system in order to fix any error, this is User will interact with the system only through the mouse following deterministic rules:

(i) Left-click over the image: mouse pointer coordinates are a correct corner of some zone in h^*.

(ii) Right-click over the image: layout is correct (ie. "OK signal")

(iii) Click outside the image: Close the system.

As consequence of use a deterministic feedback, the signal does not need to be "decoded" and it can be used directly [3] on Eq. 3.7 as the feedback f. Feedback under this context means a constrain under the search space $\mathcal{H}$, this is, rule (i) constrain the search space to only those hypotheses where the selected coordinates in the image are part of h, then, that coordinate will considered as "fixed". Also, history is important to constrain search space, to only those zone corners in the hypothesis which are not "fixed" by the user on previous iterations.

Algorithm 2: Interactive Algorithm

Data: a set of images $\mathcal{X}$, Prob site model $\mathcal{P}$, Prob layout Model $\mathcal{Q}$
Result: best, under the model, array of layout hypothesis $\hat{\boldsymbol{h}}$
for $x \in \mathcal{X}$ **do**
    $\hat{h}_x$ = Eq. 3.6;
    while $f \neq OK$ **do**
        $h' = \hat{h}_x$;
        f = decodeUserFeedbak();
        $\hat{h}_x$ = Eq. 3.7;
return $\hat{\boldsymbol{h}}$

Bibliography

[1] Cattoni, R., Coianiz, T., Messelodi, S., and Modena, C. M. (1998). Geometric layout analysis techniques for document image understanding: a review. Technical report.

[2] Duda, R. O., Hart, P. E., and Stork, D. G. (2000). *Pattern Classification (2Nd Edition)*. Wiley-Interscience.

[3] Toselli, A. H., Vidal, E., and Casacuberta, F. (2011). *Multimodal Interactive Pattern Recognition and Applications*. Springer.

CHAPTER 4

EXPERIMENTS AND RESULTS

Chapter Outline

4.1 Corpus Description

To develop and test the new method proposed in this document a manuscript under following requirements is preferred to be selected:

- Reasonably good state of preservation.
- Available in some digital format.
- Reasonably large.
- Well-defined layout zones.
- Layout should be complex enough to exemplify method strengths and weaknesses.
- Ground trough available or easy to build.

Under those constrains, the manuscript chosen for the present work is the first tome of a seven volume manuscript entitled "Historia de las Plantas", PLANTAS for short, a XVII century handwritten botanical specimen book compiled by Bernardo Cienfuegos, one of the most outstanding Spanish botanists in the XVII century. The first volume has 49 pages at the beginning comprising indices, reference tables, a botanical glossary in different languages, and a 36-page preface written by Cienfuegos. This is followed by 887 numbered pages that contain 152 chapters about cereals and related plants, including 126 botanical illustrations. All in all, the first volume has 1 035 pages, containing about 20000 handwritten text lines. This corpus is already digitized at 300ppi in 24 bit RGB color, available as JPG images along with their respective ground trough layout in PAGE XML format [7] compiled by PRHLT group [1] using seven categories, namely: catch-word, heading, marginalia, page-number, paragraph, signature-mark, float (illustrations); see Figure 4.1 for reference.

In this stage of the presented work only a sub-set of 39 pages was considered, in order to comply with stage and time restrictions. Pages that contains indice, reference tables and illustrations were excluded since goals of this stage are restricted to the main paragraph. 22 of those pages were selected for training the model, and the remaining 17 for test (see 6.1 for reference).

4.2 Implementation Notes

System was implemented mainly in Python 2.7 due the facilities provided by the language (N-dimensional array structures, image processing tools, plotting

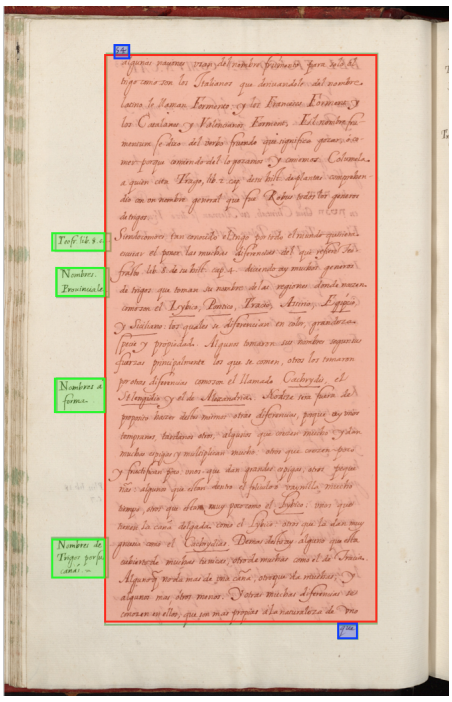


Figure 4.1: PLANTAS Layout Zones; green for marginalia, red for paragraph and blue for catch-word

tools, etc) and how quickly is possible to develop a new piece of software. Besides, CRFSuite is used to handle CRF training and tagging steps. On this chapter some implementation details will be explained for clarification; most of them are related to non-direct implementation of the theory explained on Chapters 2 and 3, and implemented towards code optimization.

4.2.1 Integral Image

Integral Image (or Summed Area Table) is a data structure and algorithm first used by Crow [3] in Computer Graphics and introduced by Viola and Jones [8] to Computer Vision it is widely used for quickly and efficiently generating the sum of values in a rectangular subset of a grid. The Integral Image $\boldsymbol{I}$ of an $M \times N$ input image A is defined as a 2D cumulative sum of A:

$$\boldsymbol{I}[x, y] = \sum_{x' \leq x, y' \leq y} A(x', y') \quad x \leq M, y \leq N \tag{4.1}$$

Then, using the integral image any rectangular sum can be computed in four array references [8] (see Figure 4.2), this is four accesses to the I matrix instead of $(x * y)$ accesses to A using the direct approach.

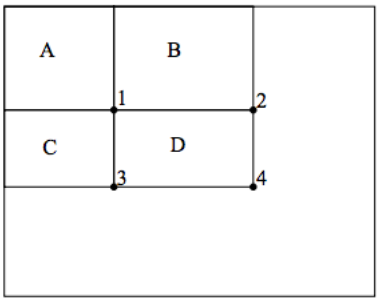


Figure 4.2: The sum of the pixels within rectangle D can be computed with four array references. The value of the Integral Image at location 1 is the sum of the pixels in rectangle A. The value at location 2 is A + B, at location 3 is A + C, and at location 4 is A + B + C + D. The sum within D can be computed as $4 + 1 - (2 + 3)$.[8]

Since CRF model provides the probability of each site in the site set S, we can see it as a matrix and use Integral Image for all related calculations. For example on Eq. 3.7 we need to compute the sum of the probability of each site inside each k layout-zone; this can be computed as:

$$\sum_{d=1}^{D_k} \log \mathrm{P}(s_d|(\boldsymbol{u}_k, \boldsymbol{b}_k)) = \boldsymbol{I}_k[\boldsymbol{b}_k] + \boldsymbol{I}_k[\boldsymbol{u}_k] - \boldsymbol{I}_k[\boldsymbol{u}_{kr}, \boldsymbol{b}_{kc}] - \boldsymbol{I}_k[\boldsymbol{b}_{kr}, \boldsymbol{u}_{kc}] \quad (4.2)$$

with vectors $\boldsymbol{u}_k$ and $\boldsymbol{b}_k$ defined in Section 3.1.

4.2.2 Element-wise to matrix evaluation

User interaction requires a system fast enough to keep real time feeling to the user, this is less than 1.0 seconds [6]. On this system user interacts as explained on Section 3.7, but direct implementation of Eq. 3.7 using `for` loops is just not fast enough to keep real time feeling ($\approx$ 25 seconds per click). In order to reduce the time consumed by the CPU to compute Eq. 3.7 a matrix-like version of the equation is implemented. This allows the system to use `numpy` multidimensional and broadcasting features to reduce delay.
Then, Eq. 3.7 is transformed to Eq. 4.3, where sums have been expanded to specific the specific case of this work (only main paragraph), $\boldsymbol{I}_k$ means for the Integral Image of the probabilities of the layout zone k, $\boldsymbol{P}_{\boldsymbol{u}_1}$ and $\boldsymbol{P}_{\boldsymbol{b}_1}$ are the probability matrix from $\boldsymbol{u}_1$ and $\boldsymbol{b}_1$ GMM models respectively, f_r and f_c are the

decoded feedback from the user as the row and column selected respectively.

$$
\begin{aligned}
\hat{h}_{\boldsymbol{b}_1} &= \arg\min \boldsymbol{I}_0[-1,-1] \\
&\quad - (\boldsymbol{I}_0[f_r:, f_c:] + \boldsymbol{I}_0[f_r-1, f_c-1] - \boldsymbol{I}_0[f_r-1, f_c:] - \boldsymbol{I}_0[f_r:, f_c-1]) \\
&\quad + (\boldsymbol{I}_1[f_r:, f_c:] + \boldsymbol{I}_1[f_r-1, f_c-1] - \boldsymbol{I}_1[f_r-1, f_c:] - \boldsymbol{I}_1[f_r:, f_c-1]) \\
&\quad + \boldsymbol{P}_{\boldsymbol{u}_1}[f_r, f_c] + \boldsymbol{P}_{\boldsymbol{b}_1}[f_r:, f_c:] \\
\hat{h}_{\boldsymbol{u}_1} &= \arg\min \boldsymbol{I}_0[-1,-1] \\
&\quad - (\boldsymbol{I}_0[f_r, f_c] + \boldsymbol{I}_0[:f_r-1, :f_c-1] - \boldsymbol{I}_0[:f_r-1, f_c] - \boldsymbol{I}_0[f_r, :f_c-1]) \\
&\quad + (\boldsymbol{I}_1[f_r, f_c] + \boldsymbol{I}_1[:f_r-1, :f_c-1] - \boldsymbol{I}_1[:f_r-1, f_c] - \boldsymbol{I}_1[f_r, :f_c-1]) \\
&\quad + \boldsymbol{P}_{\boldsymbol{u}_1}[1:f_r, 1:f_c] + \boldsymbol{P}_{\boldsymbol{b}_1}[f_r, f_c]
\end{aligned} \tag{4.3}
$$

Which is the correct notation for this kind of equation, where most of the terms are matrix

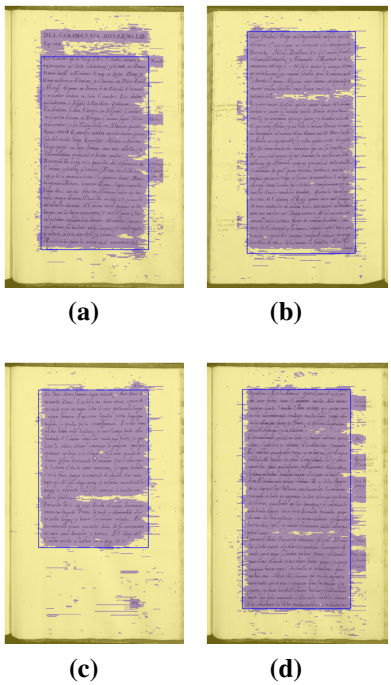
Under this approach, time to compute min value is reduced from ≈ 25 seconds to ≈ 0.06 seconds, which is enough for current application.

4.3 Experiments

Experiments have been conducted over selected corpus to obtain model parameters such as features to train CRF models, window size of those features and training algorithm, between others. Parameters search performed is not exhaustive since main goal of this stage is not to get the best model, but to demonstrate the interactive algorithm features.

4.3.1 Conditional Random Fields

CRF performance is highly dependent of features selection; because of that, some values of image zoom (Z), window size (W), grid size (G). AROW [5] training algorithm have been selected to train the CRF model because is proved to be very fast and results are similar to L-BFGS [2] or others. Results are presented on Table 6.2 for reference. In view of the results, following parameters are selected to train the final model: Z=0.3, W=33 and G=3. Quantitative results are shown in Table 4.1 and some examples of qualitative results in Figure 4.3



(a) **(b)**

(c) **(d)**

Figure 4.3: CRF's qualitative results, pixels classified as background in yellow, pixels classified as paragraph in purple. Ground through rectangle in blue. Pages a) 0944, b) 0945, c) 0948 and d) 0958.

Table 4.1: CRF's site level quantitative results.

Page	Precision	Recall	F1
0944	0.908	0.914	0.907
0945	0.953	0.953	0.953
0946	0.958	0.956	0.956
0947	0.951	0.952	0.951
0948	0.949	0.951	0.950
0956	0.894	0.896	0.891
0957	0.946	0.945	0.945
0958	0.931	0.927	0.927
0959	0.952	0.952	0.952
0960	0.955	0.954	0.954
0961	0.950	0.950	0.950
0962	0.909	0.921	0.910
0963	0.919	0.914	0.916
0964	0.909	0.910	0.908
0965	0.942	0.942	0.942
0966	0.947	0.947	0.947
0967	0.945	0.946	0.945
Average	0.936	0.936	0.936

4.3.2 Connected Components Labeling (CCL) Approach

Connected components algorithms based on morphological operations are a simple method to detect connected objects or regions in binary images. A simple version of this kind of algorithms [4] is implemented as a point of comparison for proposed method. This is, all adjacent sites classified as "paragraph" by the CRF model are grouped, then we search for the minimum rectangle where all sites of the same group fits and finally, based on user experience, only the biggest rectangle is selected. See Figure 4.4 for some qualitative examples and Table 4.2 for quantitative results.

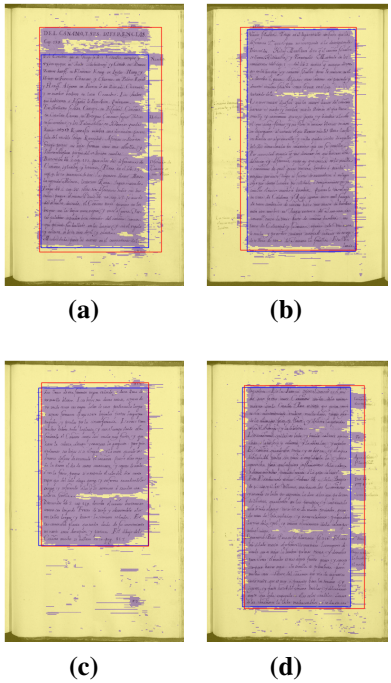


Figure 4.4: CCL results examples, red line. Ground through added for reference, blue line. Pages a) 0944, b) 0945, c) 0948 and d) 0958.

Table 4.2: CCL quantitative results.

Page	MatchScore	GoSR
0944	0.862	0.768
0945	0.958	0.926
0946	0.929	0.877
0947	0.969	0.944
0948	0.965	0.937
0956	0.873	0.782
0957	0.956	0.921
0958	0.923	0.863
0959	0.984	0.971
0960	0.932	0.879
0961	0.967	0.942
0962	0.861	0.765
0963	0.970	0.944
0964	0.878	0.787
0965	0.948	0.908
0966	0.929	0.873
0967	0.950	0.912
Average	0.933	0.882

4.3.3 Prior-Probability Approach

Prior-Probability is used to estimate the best "paragraph" coordinates, this is we maximize Eq. 3.6 over all $(\boldsymbol{u}_k, \boldsymbol{b}_k)$ in the image range using brute force approach and methods explained in Section 4.2. See Figure 4.5 for some qualitative results examples, and Table 4.3 for quantitative results.

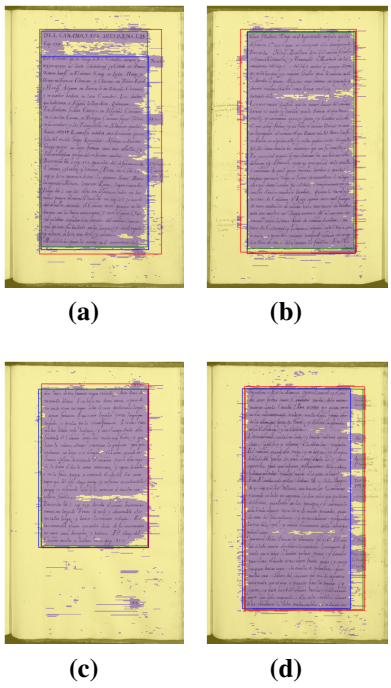


Figure 4.5: Proposed method results example, blue line for ground trough, red line for connected components labeling approach, green line for proposed approach. Pages a) 0944, b) 0945, c) 0948 and d) 0958.

Table 4.3: Proposed method quantitative results.

Page	MatchScore	GoSR
0944	0.924	0.862
0945	0.975	0.956
0946	0.988	0.977
0947	0.968	0.943
0948	0.983	0.969
0956	0.882	0.791
0957	0.978	0.960
0958	0.930	0.875
0959	0.969	0.945
0960	0.988	0.978
0961	0.973	0.952
0962	0.927	0.869
0963	0.937	0.888
0964	0.934	0.878
0965	0.969	0.944
0966	0.974	0.951
0967	0.966	0.940
Average	0.957	0.922

4.3.4 Interactive Approach

User is allowed to change system hypothesis by a simple click over the image. This feedback is decoded and new hypothesis is presented to the user. Number of clicks needed by the user to define the main paragraph and new hypothesis are recorded. Quantitative results are presented on Table 4.4 along with the number of clicks performed by an user.

Table 4.4: Post-user-feedback quantitative results.

Page	MatchScore	GoSR	# Clicks
0944	0.978	0.960	1
0945	0.975	0.956	0
0946	0.988	0.977	0
0947	0.968	0.943	0
0948	0.983	0.969	0
0956	0.974	0.952	2
0957	0.978	0.960	0
0958	0.983	0.969	1
0959	0.969	0.945	0
0960	0.988	0.978	0
0961	0.973	0.952	0
0962	0.975	0.955	1
0963	0.956	0.921	1
0964	0.971	0.947	1
0965	0.969	0.944	0
0966	0.974	0.951	0
0967	0.966	0.940	0
Average	0.975	0.954	—

4.4 Results Discussion

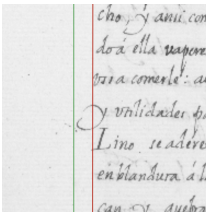
CRF model performed an average of 93.6% F1-score. Despite the set of features selected were very elemental (only site color intensity and position) and a non-exhaustive parameter search was done, results are very promising for further stages. Miss-classification of marginalia and title zones could be because color intensity is a feature more to identify text than to identify the different layout zones. That means layout zone classification lays mostly on position feature. Also only two classes have been used for training the CRF model. Although this model provides posterior probability needed for interactive model, feature selection needs to be improved for further stages of this project.

CCL is highly dependent on geometric distribution of posterior probability from CRF model and, after segmentation, there is no easy and general way to improve results. For this reason this approach is used only as a point of comparison for proposed method. Performance computed by GoSR and MatchScore are pretty different (93.2% and 83.2% respectively on average) because of the first method is most a measure of how similar are the polygons, but second method finds how much of the information is on both polygons. Notice that almost all marginalia

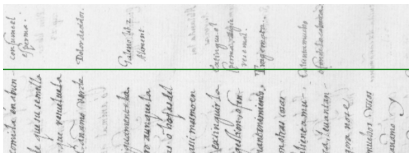
and title zones have been classified as paragraph. This is due to the high dependence of geometric distribution.

Prior-probability plays a main role in next approach, most of the marginalia zones are not longer classified as paragraph, and rectangle boundary is stretched to text boundary which relies on a average 4% GoSR improvement over CCL approach. This is up to 13% improvement on complex cases (see Table 4.3).

Finally, methods studied on bibliography and the method proposed in this work still have errors that must be fixed manually by some human. The interactive approach based on the proposed method provides the framework to help user to fix those errors. Under that premise not only the system performance must be computed, but the user effort as well. On this seventeen pages corpus, 65% of the times user made no changes to the first provided hypothesis. Thus, hypothesis is good enough to identify the paragraph. On the other hand, 30% of the times user performs only one click in order to fix errors on the hypothesis. Finally, only in one case two clicks were needed. Performance average is 97.5% and 95.3% on MathScore and GoSR methods respectively; notice that 100% was not reached in any case because users have a different concept of how much the border of the zone needs to be to the border of the text, or cases where two o more blocks cannot be divided by a single line. See examples on Figure 4.6.



(a)



(b)

Figure 4.6: Examples of different zone boundaries provided by different users. a) small section of a character is out due adjust to text border, b) single horizontal line cannot divide upper and bottom zones.

Bibliography

[1] Bosch, V., Bordes-Cabrera, I., Muñoz, P. C., Hernández-Tornero, C., Leiva, L. A., Pastor, M., Romero, V., Toselli, A. H., and Vidal, E. (2014). Computer-assisted Transcription of a Historical Botanical Specimen Book : Organization and Process Overview Categories and Subject Descriptors. In *Proceedings of the First International Conference on Digital Access to Textual Cultural Heritage*, pages 125–130, Madrid, Spain.

[2] Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM J. Sci. Comput.*, 16(Sept. 1995):1190–1208.

[3] Crow, F. C. (1984). Summed-area tables for texture mapping. *ACM SIGGRAPH Computer Graphics*, 18(3):207–212.

[4] Gonzalez, R. C. and Woods, R. E. (2008). *Digital Image Processing*. Prentice Hall, 3 edition.

[5] Kulesza, A., Crammer, K., and Dredze, M. (2009). Adaptive Regularization of Weight Vectors. *Advances in Neural Information Processing Systems*, pages 1–9.

[6] Nielsen, J. (1993). *Usability Engineering*. Morgan Kaufmann Publishers Inc.

[7] Pletschacher, S. and Antonacopoulos, A. (2010). The PAGE (Page Analysis and Ground-truth Elements) format framework. *Proceedings - International Conference on Pattern Recognition*, pages 257–260.

[8] Viola, P. and Jones, M. (2001). Robust real-time object detection. *International Journal of Computer Vision*, 57(2):137–154.

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

Chapter Outline

5.1 Conclusions[WIP]

- A new Interactive-Probabilistic methodology for Layout Analysis on ancient documents is present.
-

5.2 Future Work[WIP]

CHAPTER 6

APPENDIX

Chapter Outline

6.1 Corpus Distribution Table

Table 6.1: Corpus Pages Distribution

Name	ID	used to
Mss_003357_0944_pag-811[843]	0944	test
Mss_003357_0945_pag-812[844]	0945	test
Mss_003357_0946_pag-813[845]	0946	test
Mss_003357_0947_pag-814[846]	0947	test
Mss_003357_0948_pag-815[847]	0948	test
Mss_003357_0956_pag-823[855]	0956	test
Mss_003357_0957_pag-824[856]	0957	test
Mss_003357_0958_pag-825[857]	0958	test
Mss_003357_0959_pag-826[858]	0959	test
Mss_003357_0960_pag-827[859]	0960	test
Mss_003357_0961_pag-828[860]	0961	test
Mss_003357_0962_pag-829[861]	0962	test
Mss_003357_0963_pag-830[862]	0963	test
Mss_003357_0964_pag-831[863]	0964	test
Mss_003357_0965_pag-832[864]	0965	test
Mss_003357_0966_pag-833[865]	0966	test
Mss_003357_0967_pag-834[866]	0967	test
Mss_003357_0158_pag-053[057]	0158	train
Mss_003357_0159_pag-054[058]	0159	train
Mss_003357_0161_pag-056[060]	0161	train
Mss_003357_0162_pag-057[061]	0162	train
Mss_003357_0163_pag-058[062]	0163	train
Mss_003357_0176_pag-071[075]	0176	train
Mss_003357_0177_pag-072[076]	0177	train
Mss_003357_0178_pag-073[077]	0178	train
Mss_003357_0179_pag-074[078]	0179	train
Mss_003357_0180_pag-075[079]	0180	train
Mss_003357_0181_pag-076[080]	0181	train
Mss_003357_0182_pag-077[081]	0182	train
Mss_003357_0183_pag-078[082]	0183	train
Mss_003357_0184_pag-079[083]	0184	train
Mss_003357_0185_pag-080[084]	0185	train
Mss_003357_0186_pag-081[085]	0186	train
Mss_003357_0187_pag-082[086]	0187	train
Mss_003357_0188_pag-083[087]	0188	train
Mss_003357_0189_pag-084[088]	0189	train
Mss_003357_0190_pag-085[089]	0190	train
Mss_003357_0191_pag-086[090]	0191	train
Mss_003357_0192_pag-087[091]	0192	train

6.2 CRF's parameter search.

Table 6.2: CRF's parameters search results. Z=zoom, W=window, G=granularity, FT= feature extraction time, TrT= Training time, TeT= Test time, P= Precision, R= Recall, F1= F1-score.

	Z	W	G	FT[s]	TrT[s]	TeT[s]	P	R	F1
1	0.1	9	3	4.916	16.822	1.150	0.899	0.900	0.897
2	0.1	9	9	1.151	0.074	0.117	0.868	0.865	0.859
3	0.1	9	12	0.937	0.035	0.062	0.878	0.874	0.872
4	0.1	17	3	20.448	16.054	0.962	0.894	0.884	0.882
5	0.1	17	9	3.619	0.074	0.092	0.895	0.886	0.884
6	0.1	17	12	2.727	0.019	0.052	0.852	0.833	0.828
7	0.1	33	3	23.603	0.997	0.792	0.843	0.811	0.817
8	0.1	33	9	3.920	0.044	0.084	0.858	0.796	0.804
9	0.1	33	12	2.815	0.015	0.043	0.850	0.714	0.714
10	0.1	65	3	32.006	1.050	0.704	0.684	0.659	0.668
11	0.1	65	9	5.078	0.064	0.067	0.820	0.667	0.687
12	0.1	65	12	3.528	0.029	0.037	0.863	0.712	0.742
13	0.2	9	3	18.714	142.845	4.559	0.866	0.868	0.866
14	0.2	9	9	2.714	0.825	0.523	0.852	0.852	0.847
15	0.2	9	12	2.127	0.290	0.261	0.882	0.885	0.882
16	0.2	17	3	88.976	122.980	4.235	0.866	0.863	0.858
17	0.2	17	9	11.068	5.450	0.422	0.863	0.858	0.851
18	0.2	17	12	6.758	0.219	0.225	0.895	0.896	0.892
19	0.2	33	3	105.599	81.756	4.396	0.905	0.901	0.901
20	0.2	33	9	13.263	1.048	0.410	0.895	0.888	0.887
21	0.2	33	12	8.095	0.246	0.214	0.896	0.889	0.888
22	0.2	65	3	173.887	76.256	3.759	0.853	0.848	0.850
23	0.2	65	9	20.629	0.477	0.393	0.837	0.812	0.819
24	0.2	65	12	12.392	0.287	0.220	0.829	0.803	0.808
25	0.3	9	3	41.491	457.521	12.270	0.829	0.832	0.829
26	0.3	9	9	5.174	19.041	1.207	0.859	0.860	0.859
27	0.3	9	12	3.271	1.414	0.554	0.866	0.868	0.867
28	0.3	17	3	195.334	341.149	10.868	0.881	0.882	0.880
29	0.3	17	9	23.479	17.754	1.117	0.880	0.883	0.880
30	0.3	17	12	13.886	2.578	0.563	0.885	0.888	0.884
31	0.3	33	3	24.153	20.972	9.582	0.936	0.937	0.936
32	0.3	33	9	29.088	15.112	0.973	0.898	0.898	0.896
33	0.3	33	12	16.788	1.933	0.527	0.895	0.892	0.890
34	0.3	65	3	396.537	190.954	9.099	0.901	0.898	0.899
35	0.3	65	9	45.363	2.987	0.927	0.891	0.881	0.883
36	0.3	65	12	26.538	6.580	0.502	0.881	0.870	0.870

NOMENCLATURE

AROW	Adaptive Regularization of Weights
BFGS	Broyden–Fletcher–Goldfarb–Shanno
CCL	Connected Components Labeling
CRF	Conditional Random Field
DLA	Document Layout Analysis
GMM	Gaussian Mixture Model
IPR	Interactive Pattern Recognition
L-BFGS	Limited-memory BFGS
PCA	Principal Components Analysis