



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

# Interactive Layout Analysis

---

IARFID Master Thesis

Lorenzo Quirós Díaz

1. Introduction
  - 1.1 Problem Definition
  - 1.2 Related work
2. Proposed Method
  - 2.1 Fundaments
3. Experiments
  - 3.1 Evaluation Measures
  - 3.2 Corpus Overview
  - 3.3 Results
4. Conclusions and Future work

# Outline

- 1. Introduction
  - 1.1 Problem Definition
  - 1.2 Related work
- 2. Proposed Method
  - 2.1 Fundaments
- 3. Experiments
  - 3.1 Evaluation Measures
  - 3.2 Corpus Overview
  - 3.3 Results
- 4. Conclusions and Future work

# Introduction

El mamo Galeno en el libro de simpliciam dize ansi. La  
simiente del Canamo quita de tal manera los flujos y  
los tumores que si se come demasiado desca el estomago.  
Algunos en que del zumo de la yerba Verde vision de  
ratos quitanos de la ojiva el qual quita los dolores de oido,  
y aquellos causados de obstrucion.

Hinc lib. 20. cap. 23. respite ansi mismo lo que en el dho  
capitulo pone las palabras. Los canamones quitan la virilidad  
genital en los hombres. Su zumo saca qualquier quiconillo  
o animal que entre en el oido por donde dolor de cabeza. Es-

## Document Transcription

Is an easy task for human beings, but hard problem for computers.

# HTR Process for Ancient Documents

## Problem Definition

- An image  $\mathcal{X} = \{x_{1,1}, x_{1,2}, \dots, x_{n,m}\}$ , which is associated with a rectangular grid  $G$  of size  $n \times m$ .

## Problem Definition

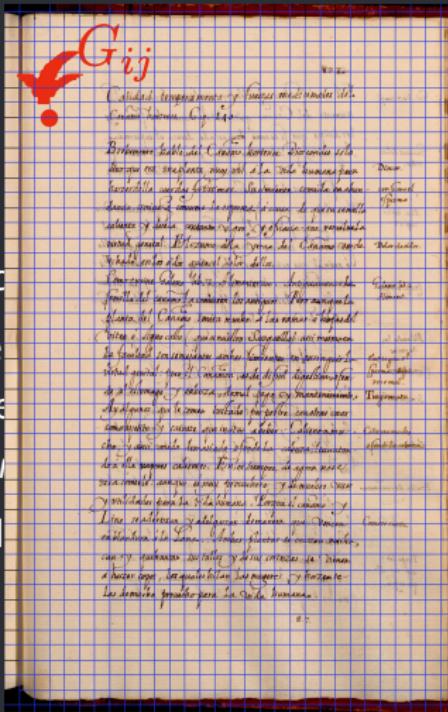
- An image  $\mathcal{X} = \{x_{1,1}, x_{1,2}, \dots, x_{n,m}\}$ , which is associated with a rectangular grid  $G$  of size  $n \times m$ .
- Each image site  $s$  is associated to a cell in the grid defined by its coordinates over  $G$  and denoted  $G_{ij}$ ,  $1 \leq i \leq n$ ,  $1 \leq j \leq m$ .

## Problem Definition

- An image  $\mathcal{X} = \{x_{1,1}, x_{1,2}, \dots, x_{n,m}\}$ , which is associated with a rectangular grid  $G$  of size  $n \times m$ .
- Each image site  $s$  is associated to a cell in the grid defined by its coordinates over  $G$  and denoted  $G_{ij}$ ,  $1 \leq i \leq n$ ,  $1 \leq j \leq m$ .
- The site set is denoted  $S = \{s_1, s_2, \dots, s_D\}$   $1 \leq D \leq n \times m$ .

# Problem Definition

- An image  $\mathcal{X}$  = rectangular grid
- Each image site coordinates over  $G_{ij}, \quad 1 \leq i \leq n, 1 \leq j \leq m$
- The site set is defined by



The site set is associated with a grid defined by its dimensions  $n \times m$ .  
 $\} \quad 1 \leq D \leq n \times m.$

## Problem Definition [Cont...]

- Now lets define  $L = \{l_1, l_2, \dots, l_\ell\}$  the set of all the possible zones in a Layout.

## Problem Definition [Cont...]

- Now lets define  $L = \{l_1, l_2, \dots, l_\ell\}$  the set of all the possible zones in a Layout.
- Then, define a *structured hypotheses space* [7]  
 $\mathcal{H} = \{h^1, h^2, \dots, h^T\}$  over the site set  $S$ , where  
 $h^t \subseteq L, \quad 1 \leq t \leq T$

## Problem Definition [Cont...]

- Now lets define  $L = \{l_1, l_2, \dots, l_\ell\}$  the set of all the possible zones in a Layout.
- Then, define a *structured hypotheses space* [7]  
 $\mathcal{H} = \{h^1, h^2, \dots, h^T\}$  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.

## Problem Definition [Cont...]

- Now lets define zones in a Layout
  - Then, define a set of hypotheses  

$$\mathcal{H} = \{h^1, h^2, \dots\}$$

$$h^t \subseteq L, \quad 1 \leq t \leq T$$
  - We want the hypothesis to be consistent with the site set.

<span style="border: 1px solid green; border-radius: 50%; padding: 2px 5px; display: inline-block;">Número 2</span> <span style="border: 1px solid green; border-radius: 50%; padding: 2px 5px; display: inline-block;">Número 3</span> <span style="border: 1px solid green; border-radius: 50%; padding: 2px 5px; display: inline-block;">Número 4</span> <span style="border: 1px solid green; border-radius: 50%; padding: 2px 5px; display: inline-block;">Número 5</span>	<p>que pague dentro de los diez años, pena de 200 mil reales con las Indias que devendrá del presidente de la Real Fábrica y otras fábricas. Los demás y los Cuadros y Valores de la Caja, devendrán en sucesión de los del año pasado que no se paguen dentro de los diez años del presidente de la Real Fábrica. Caballos a pie que no tengan el valor de 100 pesos cada uno. Paseo, 16 de mayo de 1787. Los que no paguen dentro de los diez años, se quitarán de la suerte general que fuere. Redactó todos los presentes de acuerdo.</p> <p>Sedentaria vez consulta D. José de Villa, Alcalde mayor de la Ciudad de México, del que ejerce la facultad de dar la orden de que se despidan los medios generales de trigo con ración de veinte y tres gramos cada media onza, en el Lugar, Precio, Tiempo, Año, Legua y Distancia, de donde se determinará en todo procedimiento que se requiera. Algunos tienen sus medios propios fijados particularmente. Lo que a mí concierne, son los trigos que por su diferencia en el llamado Cuchimilco, el Tlaxcalteca y el Moreliano. Andén sea hecho de papas dentro de la ración de veinte y tres gramos, papas que no se impone, también ésta, alquanto que viene más tarde, trigo que se impone y moliéndose, como que viene más tarde. Y quedando lo que no sea de los grandes espigones, una papilla que viene de la ración de veinte y tres gramos, moliéndose, como que viene, para servir de pan. Paseo, 16 de mayo de 1787. Los que no paguen dentro de los diez años, se quitarán de la suerte general que fuere. Redactó todos los presentes de acuerdo.</p> <p>Número 2 Dijo por mí -</p>
--	---

of all the possible

[7]

the best layout for

## Related work

Most methods currently developed [8, 4, 6, 5, 1, 3, 2] follow 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 in the previous step.
- (v) Clean the result (filter out small blocks, remove unsuitable blocks, etc.).

## Related work [Cont...]

### Main Characteristics

- Labeled-data are not required
- Dependent of binarization quality
- Used heuristics limit the generalization of the methods
- Designed as user-free systems

# Outline

- 1. Introduction
  - 1.1 Problem Definition
  - 1.2 Related work
- 2. Proposed Method
  - 2.1 Fundaments
- 3. Experiments
  - 3.1 Evaluation Measures
  - 3.2 Corpus Overview
  - 3.3 Results
- 4. Conclusions and Future work

# Overview

- Probabilistic model instead of Computer Vision methods

$$\hat{h} = \arg \max_{h \in \mathcal{H}} \frac{P(S|h) P(h)}{P(S)} = \arg \max_{h \in \mathcal{H}} P(h) P(S|h) \quad (1)$$

# Overview

- Probabilistic model instead of Computer Vision methods

$$\hat{h} = \arg \max_{h \in \mathcal{H}} \frac{P(S|h) P(h)}{P(S)} = \arg \max_{h \in \mathcal{H}} P(h) P(S|h) \quad (1)$$

- Conditional Random Fields (CRF) to model the conditional distribution of the pixels in each layout zone

# Overview

- Probabilistic model instead of Computer Vision methods

$$\hat{h} = \arg \max_{h \in \mathcal{H}} \frac{P(S|h) P(h)}{P(S)} = \arg \max_{h \in \mathcal{H}} P(h) P(S|h) \quad (1)$$

- Conditional Random Fields (CRF) to model the conditional distribution of the pixels in each layout zone
- Multivariable Gaussian mixture model to learn prior-probability of the layout

# Overview

- Probabilistic model instead of Computer Vision methods

$$\hat{h} = \arg \max_{h \in \mathcal{H}} \frac{P(S|h) P(h)}{P(S)} = \arg \max_{h \in \mathcal{H}} P(h) P(S|h) \quad (1)$$

- Conditional Random Fields (CRF) to model the conditional distribution of the pixels in each layout zone
- Multivariable Gaussian mixture model to learn prior-probability of the layout
- Interactive Pattern Recognition (IPR) framework to take advantage of user interaction

## CRFs

CRFs are a probabilistic framework for labeling and segmenting structured data, such as sequences, trees and lattices.

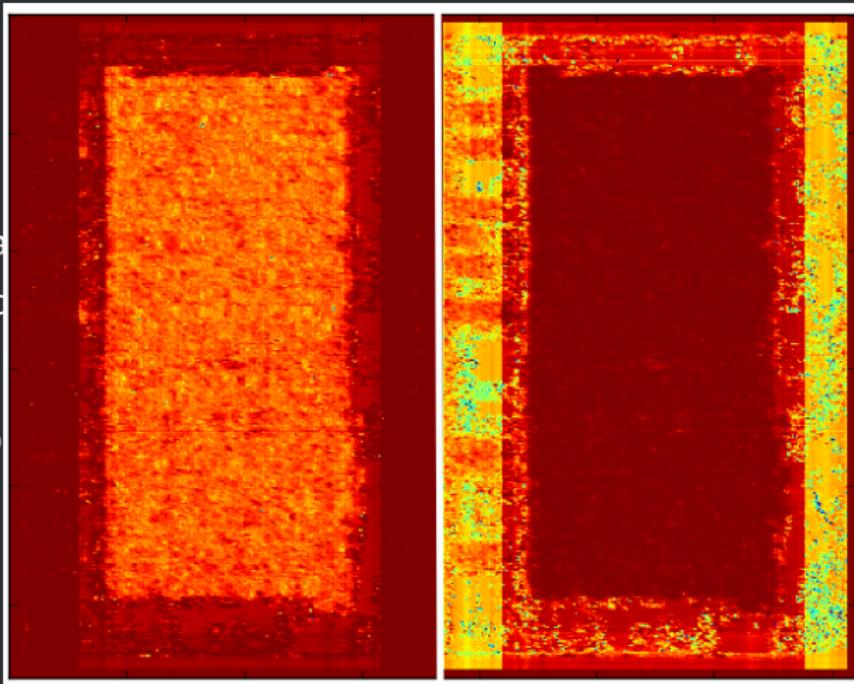
$$P(y|x) = \frac{1}{Z(x)} \prod_{\psi_d \in F} \exp \left\{ \sum_{\zeta=1}^{\mathbb{K}_d} \theta_{d\zeta} \varphi_{d\zeta}(y_d, x_d) \right\} \quad (2)$$

# CRFs

CRFs are a  
structured

P

(2)



# GMM

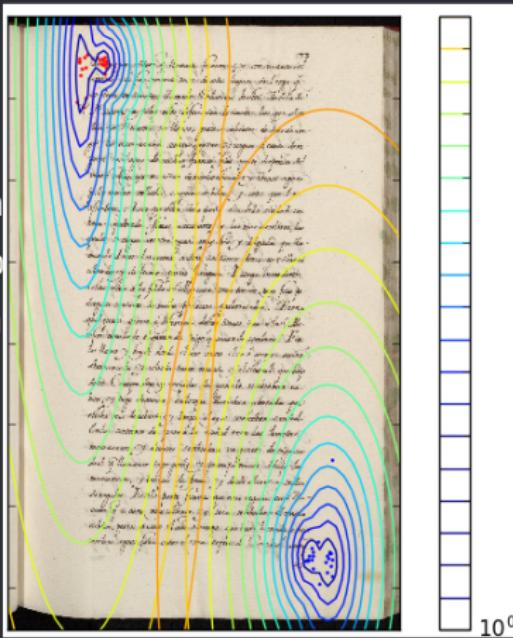
Gaussian mixture model is a probabilistic model for representing the presence of subpopulations within an overall population, by means of a set of Gaussian distributions.

$$P(h) \approx P(u) P(b) = \sum_{i=1}^M \pi_i \mathcal{N}(h|\mu_i, \Sigma_i) \quad (3)$$

# GMM

Gaussian mixture models  
presence of subpopulations  
of a set of Gaussian

$$P(h)$$

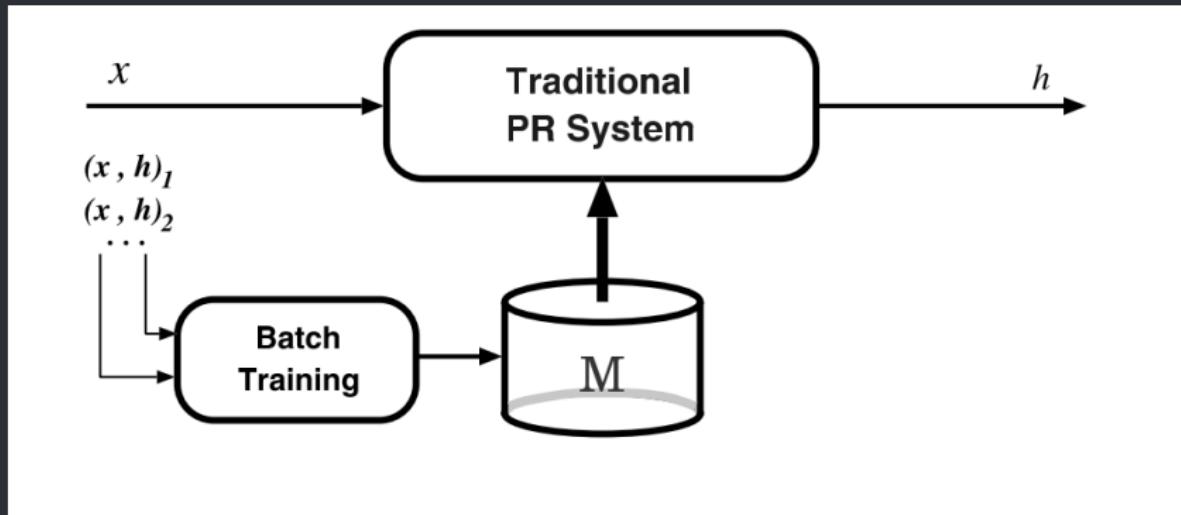


$$\Sigma_i)$$

(3)

# Classical Pattern Recognition (PR)

In classical PR,  $x$  is an input stimulus, observation or signal and  $h$  is a hypothesis or output, which the system derives from  $x$ .

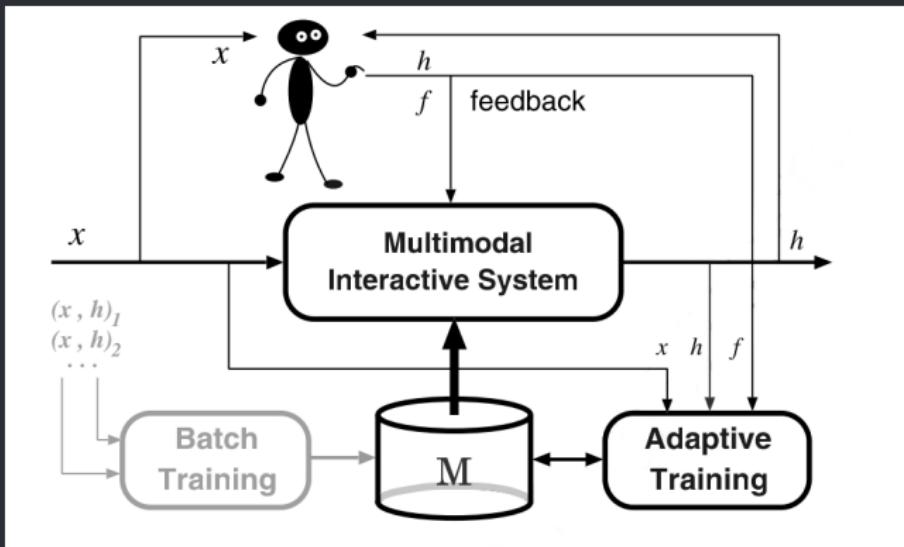


Credit: *Toselli et al. 2011*

# Interactive Pattern Recognition (IPR)

$\mathcal{M}$  model is obtained in "batch mode", as in traditional PR.

Now, during the interactive operation, the valuable user feedback signals produced in each interaction step are advantageously used.



Credit: Toselli et al. 2011

# Interactive Layout Analysis

Interactive Layout Analysis model aim to improve classical approach by taking into account user feedback:

$$\hat{h} = \arg \max_{h \in \mathcal{H}} P(h) P(S|h) \quad (4)$$



$$\hat{h} = \arg \max_{h \in \mathcal{H}} P(h, h', f) P(S|h', f, h) \quad (5)$$

# Interactive Layout Analysis

Interactive Layout Analysis model aim to improve classical approach by taking into account user feedback:

$$\hat{h} = \arg \max_{h \in \mathcal{H}} P(h) P(S|h) \quad (4)$$



$$\hat{h} = \arg \max_{h \in \mathcal{H}} P(h, h', f) P(S|h', f, h) \quad (5)$$

Joining CRF, GMM and IPR models and using  $\log$ , we get:

$$\begin{aligned} \log \hat{h} \approx \arg \max_{h \in \mathcal{H}} \sum_{k=1}^K & \left( \log \underset{\mathcal{M}_g}{P}(u_k|h', f) + \log \underset{\mathcal{M}_g}{P}(b_k|h', f) \right. \\ & \left. + \sum_{d=1}^{D_k} \log P(s_d|(u_k, b_k), h', f) \right) \quad (6) \end{aligned}$$

# Outline

- 1. Introduction
  - 1.1 Problem Definition
  - 1.2 Related work
- 2. Proposed Method
  - 2.1 Fundaments
- 3. Experiments
  - 3.1 Evaluation Measures
  - 3.2 Corpus Overview
  - 3.3 Results
- 4. Conclusions and Future work

# Evaluation Measures

- Pixel-wise F1

$$P = \frac{TP}{TP + FP} \quad (7)$$

$$R = \frac{TP}{TP + FN} \quad (8)$$

$$F1 = \frac{2 \cdot P \cdot R}{P + R} \quad (9)$$

- MatchScore

$$\text{MatchScore}(i, k) = \frac{T(h_i^* \cap h_k)}{T(h_i^* \cup h_k)} \quad (10)$$

$$FM_{i,k} = \frac{2T(h_i^* \cap h_k)}{T(h_i^*) + T(h_k)} \quad (11)$$

- Goal-Oriented Performance

$$I_{ik} = \begin{cases} h_i^* \cap h_k & \text{if } h_i^* \cap h_k \neq \emptyset \text{ and } \frac{T(I_{ik})}{T(h_k)} > T_a \\ \emptyset & \text{otherwise} \end{cases} \quad (12)$$

$$SR = \frac{\sum_{i=1}^{|h^*|} \sum_{k=1}^K w_{ik} \times T(I_{ik})}{\sum_{i=1}^{|h^*|} T(h_i^*)} \quad (13)$$

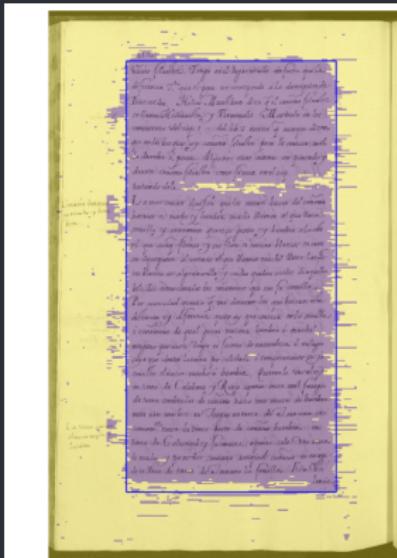
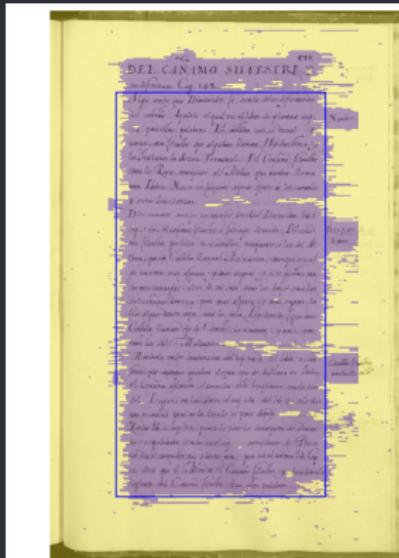
# Corpus Overview

- Manuscript entitled "Historia de las Plantas" by Bernardo Cienfuegos (XVII century)
- First volume has 1 035 pages, along with their respective ground-truth layout in PAGE XML format
- Divided in seven categories, namely: catch-word, heading, marginalia, page-number, paragraph, signature-mark, and float (illustrations)

## On this work

- Training: 22 pages
- Test: 17 pages
- Pages can contains any of the above categories, but illustrations were excluded

# Conditional Random Fields Results



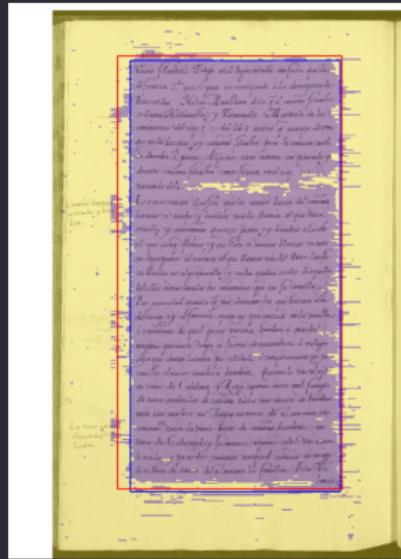
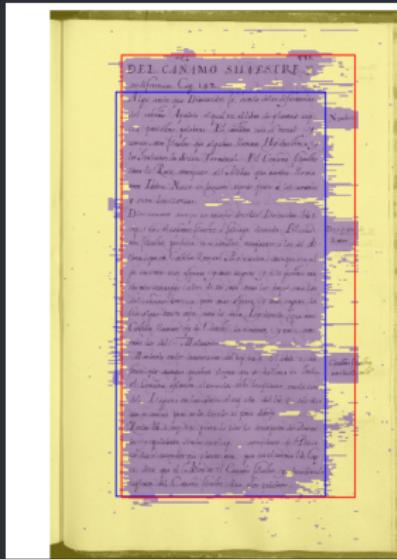
## Average Results

$$P = 0.936$$

$$R = 0.936$$

F1 = 0.936 [-1.68%, +8.63%]

# Connected Components Labeling Results

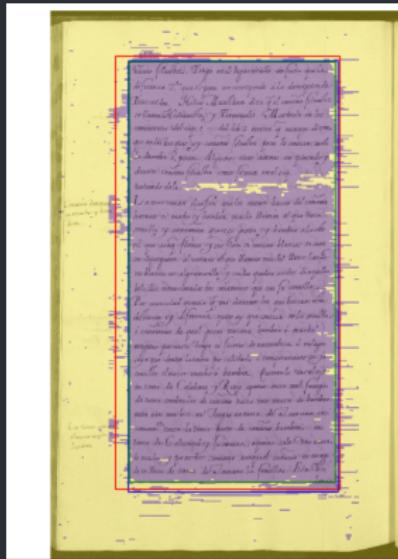
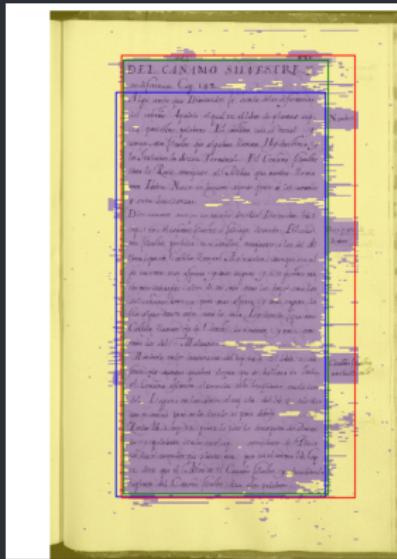


## Average Results

MatchScore = 0.933

GoSR = 0.882 [-13.23%, +10.06%]

## Probabilistic Model Results



## Average Results

MatchScore = 0.957

GoSR = 0.922 [-14.23%, +6.05%]

# DEMO

# Interactive Layout Analysis Results

## Average Results

- MatchScore: 0.975
- GoSR: 0.954
- 0 click: 11 Pages
- 1 click: 5 Pages
- 2 click: 1 Page

# Outline

- 1. Introduction
  - 1.1 Problem Definition
  - 1.2 Related work
- 2. Proposed Method
  - 2.1 Fundaments
- 3. Experiments
  - 3.1 Evaluation Measures
  - 3.2 Corpus Overview
  - 3.3 Results
- 4. Conclusions and Future work

# Conclusions

- A new method for Layout Analysis for ancient documents is presented. It was also shown that the method works, at least, to extract the main paragraph of the page.
- Inclusion of the prior-probability in the model shown a direct improvement over the CCL method, without any heuristics.
- The interactive approach provides to the user the ability to fix any error produced in the classification stage.
- HTR systems could now take into account syntactic differences between different layout-zones.
- The new method decrements the potential sources of error.

## Future work

- Remove single zone constrain
- Improve CRF's features
- Include non-deterministic feedback
- Replace brute force algorithm
- More experiments on complex corpora

# Publications

"Interactive Layout Detection" paper have been submitted to  
**IbPRIA-2017 Conference**

Notification of acceptance: January 21, 2017

# Bibliography I

- [1] A. Antonacopoulos et al. “2015\_ICDAR2015 Competition on Recognition of Documents with Complex Layouts -RDCL2015”. In: *Document Analysis and Recognition (ICDAR)*, 2015 13th International Conference on. 2015, pp. 1151–1155. ISBN: 9781479918058. DOI: [10.1109/ICDAR.2015.7333941](https://doi.org/10.1109/ICDAR.2015.7333941).
- [2] A. Antonacopoulos et al. “Historical document layout analysis competition”. In: *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR* (2011), pp. 1516–1520. ISSN: 15205363. DOI: [10.1109/ICDAR.2011.301](https://doi.org/10.1109/ICDAR.2011.301).

## Bibliography II

- [3] A. Antonacopoulos et al. “ICDAR 2013 competition on historical book recognition (HBR 2013)”. In: *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR* (2013), pp. 1459–1463. ISSN: 15205363. DOI: [10.1109/ICDAR.2013.294](https://doi.org/10.1109/ICDAR.2013.294).
- [4] Syed Saqib Bukhari et al. “Layout analysis for Arabic historical document images using machine learning”. In: *Proceedings - International Workshop on Frontiers in Handwriting Recognition, IWFHR* (2012), pp. 639–644. ISSN: 15505235. DOI: [10.1109/ICFHR.2012.227](https://doi.org/10.1109/ICFHR.2012.227).

## Bibliography III

- [5] Guillaume Lazzara, Thierry Geraud, and Roland Levillain. "Planting, growing, and pruning trees: Connected filters applied to document image analysis". In: *Proceedings - 11th IAPR International Workshop on Document Analysis Systems, DAS 2014* (2014), pp. 36–40. DOI: [10.1109/DAS.2014.36](https://doi.org/10.1109/DAS.2014.36).
- [6] Ray Smith. "Hybrid Page Layout Analysis via Tab-Stop Detection". In: *Proceedings of the 2009 10th International Conference on Document Analysis and Recognition*. IEEE Computer Society, 2009, pp. 241–245. DOI: [10.1109/ICDAR.2009.257](https://doi.org/10.1109/ICDAR.2009.257).

## Bibliography IV

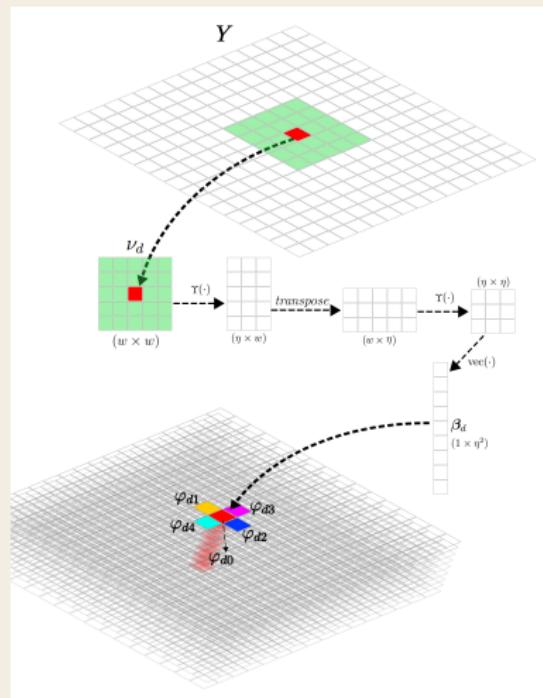
- [7] Alejandro Héctor Toselli, Enrique Vidal, and Francisco Casacuberta. *Multimodal Interactive Pattern Recognition and Applications*. Springer, 2011, p. 281. ISBN: 9780857294784. DOI: [10.1007/978-0-85729-479-1](https://doi.org/10.1007/978-0-85729-479-1).
- [8] A. M. Vil'kin, I. V. Safonov, and M. A. Egorova. "Algorithm for Segmentation of Documents Based on Texture Features". In: *Pattern Recognit. Image Anal.* 23.March, 2013 (2013), pp. 153–159. DOI: [10.1134/S1054661813010136](https://doi.org/10.1134/S1054661813010136). URL: <http://dx.doi.org/10.1134/S1054661813010136>.

# BACKUP

# CRFs Features

$$\beta_d = \text{vec}(\Upsilon(\Upsilon(\nu_d, \eta)^T, \eta))^T \quad (14)$$

- 5 related to color of site  
 $(\varphi_{d0} = \beta_d)$
- 4 related to neighborhood structure  
 $(\varphi_{d5} = \beta_{d-1} | \beta_d)$
- 2 related to site position  
 $(\varphi_{d9} = \lfloor \frac{d}{c} \rfloor)$



## Integral Image

$$\sum_{d=1}^{D_k} \log P(s_d | (u_k, b_k)) = I_k[b_k] + I_k[u_k] - I_k[u_{kr}, b_{kc}] - I_k[b_{kr}, u_{kc}] \quad (15)$$

## Matrix operations

$$\hat{h}_{b1} = \arg \min I_0[n, m]$$

$$\begin{aligned} & - (I_0[f_r : n, f_c : m] + I_0[f_r - 1, f_c - 1] - I_0[f_r - 1, f_c : m] - I_0[f_r : n, f_c - 1]) \\ & + (I_1[f_r : n, f_c : m] + I_1[f_r - 1, f_c - 1] - I_1[f_r - 1, f_c : m] - I_1[f_r : n, f_c - 1]) \\ & + P_{u1}[f_r, f_c] + P_{b1}[f_r : n, f_c : m] \end{aligned}$$

$$\hat{h}_{u1} = \arg \min I_0[n, m]$$

$$\begin{aligned} & - (I_0[f_r, f_c] + I_0[0 : f_r - 1, 0 : f_c - 1] - I_0[0 : f_r - 1, f_c] - I_0[f_r, 0 : f_c - 1]) \\ & + (I_1[f_r, f_c] + I_1[0 : f_r - 1, 0 : f_c - 1] - I_1[0 : f_r - 1, f_c] - I_1[f_r, 0 : f_c - 1]) \\ & + P_{u1}[1 : f_r, 1 : f_c] + P_{b1}[f_r, f_c] \end{aligned} \tag{16}$$