UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

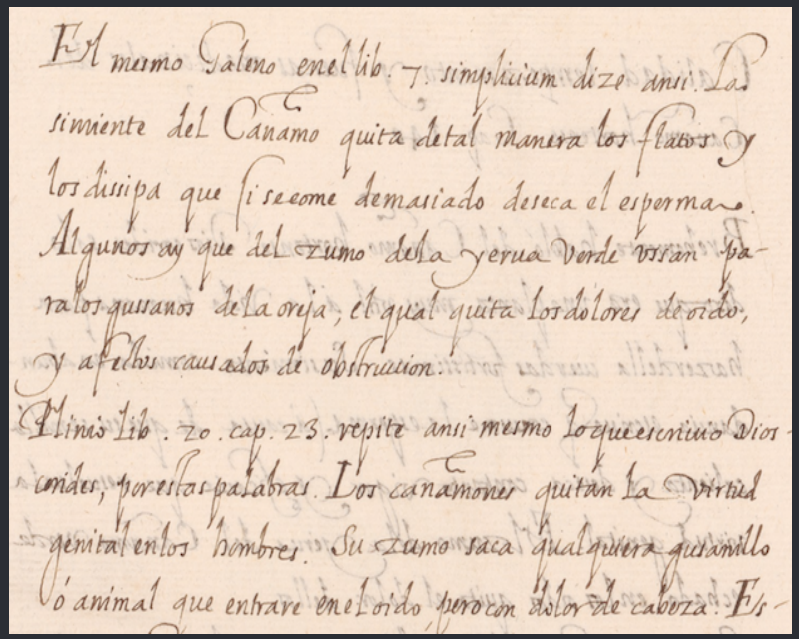
Interactive Layout Analysis

IARFID Master Thesis

Lorenzo Quirós Díaz

Outline

Introduction

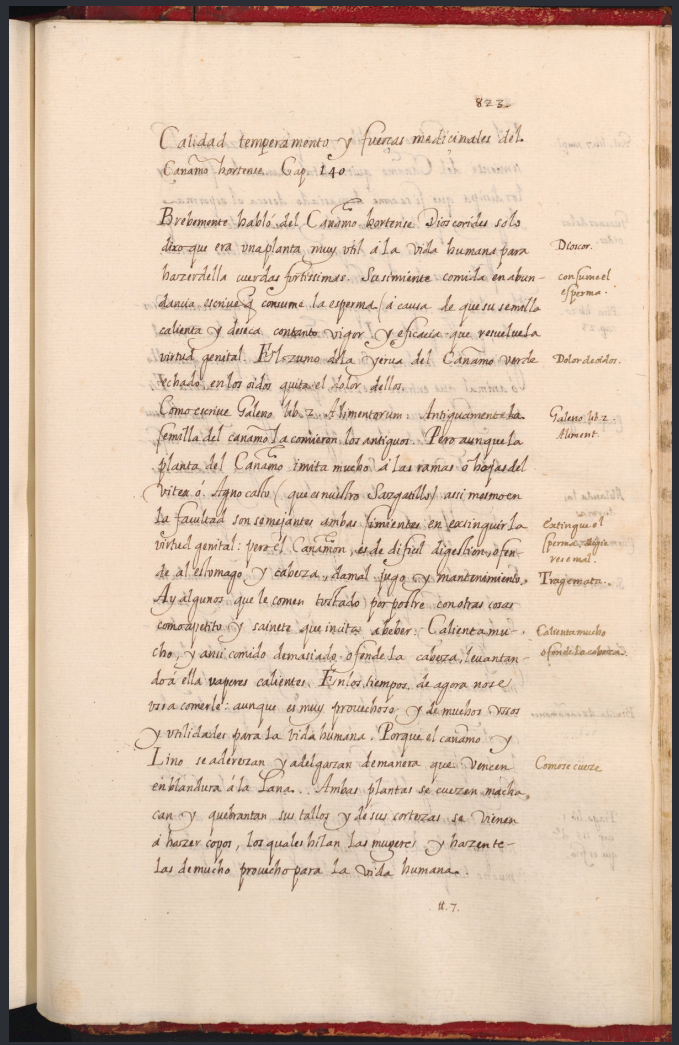

El mesmo Galeno enel lib. 7. simplicium dize assi. Los simiente del Cánamo quita detal manera los flatos y los disipa que si se come demasiado deseca el esperma. Algunos ay que del çumo dela yerua verde vsan para los gusanos dela oreja, el qual quita los dolores de oído, y a ellos causados de obstruction.

Plinio lib. 20. cap. 23. repite assi mesmo lo que escriuio Dioscorides, por estas palabras. Los canamones quitan la virtud genital en los hombres. Su çumo saca qualquiera gusanillo ó animal que entrare en el oído pero con dolor de cabeça. Es

Document Transcription

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

HTR Process for Ancient Documents



Universitat Politècnica de València

Lorenzo Quirós Díaz

Problem Definition

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

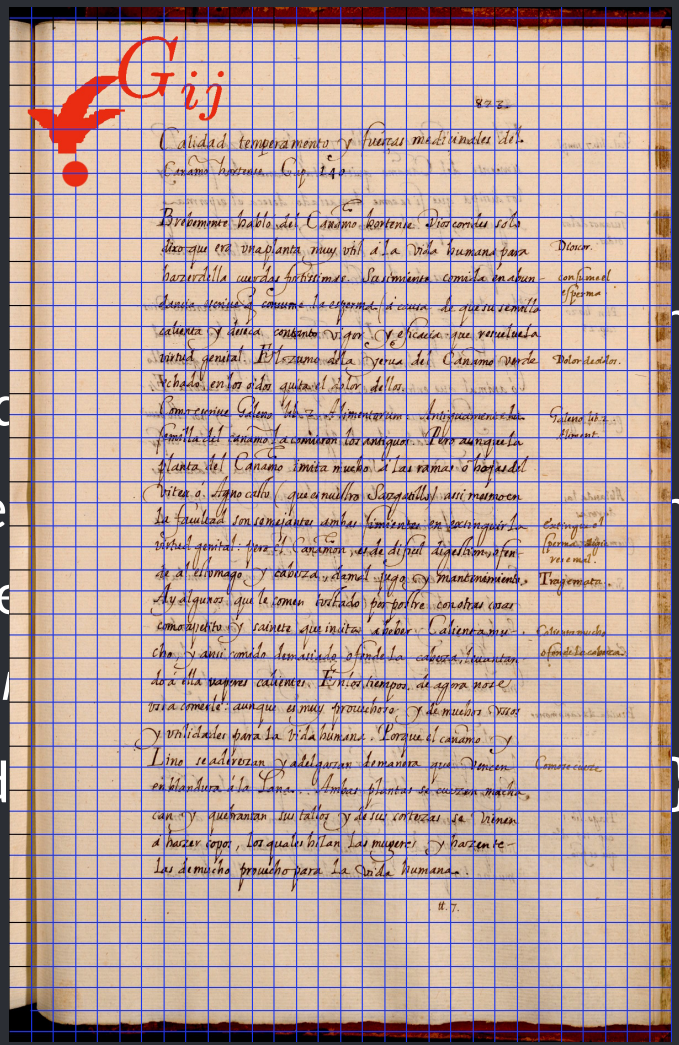
Problem Definition

- An image $\mathcal{X} = \{x_{1,1}, x_{1,2}, .., x_{n,m}\}$, 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
  $G_{ij}, \quad 1 \leq i \leq n, \quad 1 \leq j \leq m.$

Problem Definition

- An image $\mathcal{X} = \{x_{1,1}, x_{1,2}, .., x_{n,m}\}$, 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
  $G_{ij}, \quad 1 \leq i \leq n, \quad 1 \leq j \leq m.$
- The site set is denoted $S = \{s_1, s_2, ..., s_D\} \quad 1 \leq D \leq n \times m.$

Problem Definition



- An image $\mathcal{X} =$ n is associated with a rectangular gric
- Each image site ne grid defined by its coordinates ove G_{ij}, $1 \leq i \leq$
- The site set is d } $1 \leq D \leq n \times m$.

Problem Definition [Cont...]

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

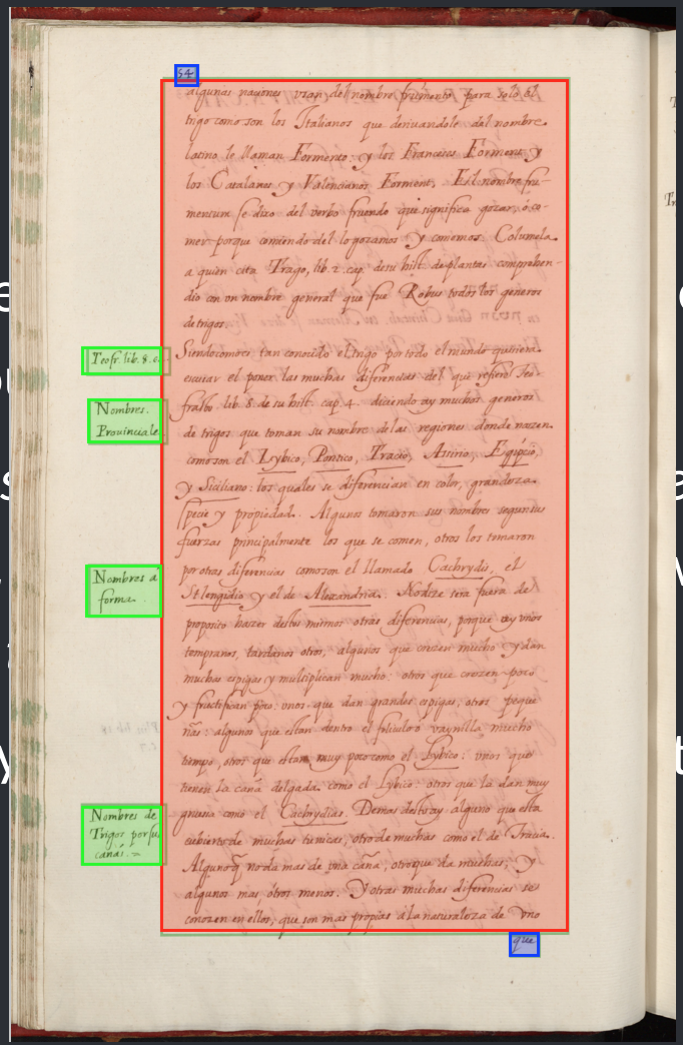
Problem Definition [Cont...]

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

Problem Definition [Cont...]

- Now lets define $L = \{l_1, l_2, \ldots, l_\ell\}$ the set of all the possible zones in a Layout.
- Then, define a *structured hypotheses space* [7]
  $\mathcal{H} = \{h^1, h^2, \ldots, h^T\}$ 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 … of all the possible zones in a Layo…
- Then, define a s… [7]
  $\mathcal{H} = \{h^1, h^2, .$… where
  $h^t \subseteq L, \quad 1 \leq$…
- We want the hy… the best layout for the site set.

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

Overview

- Probabilistic model instead of Computer Vision methods

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

Overview

- Probabilistic model instead of Computer Vision methods

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

- 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

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

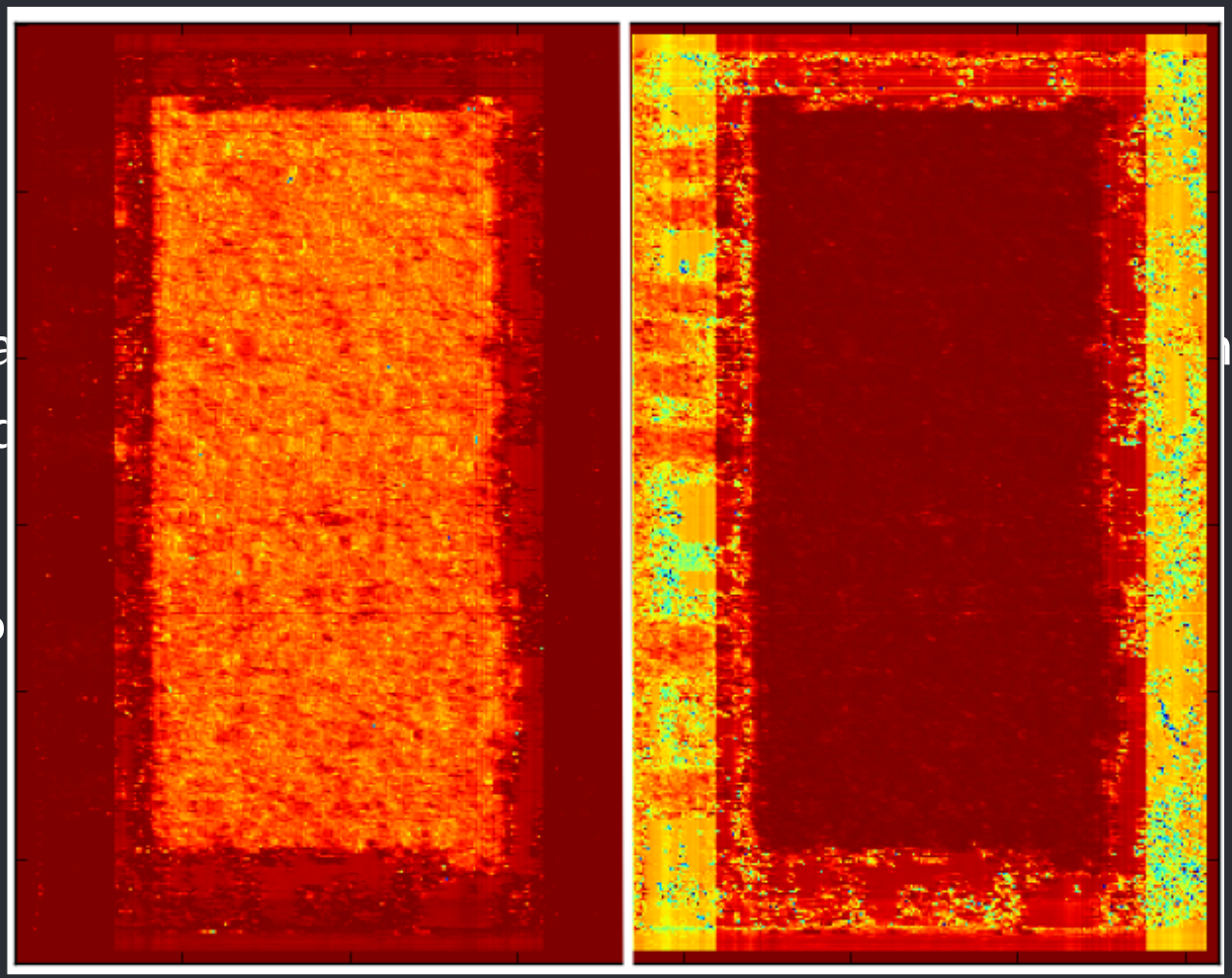
- Conditional Random Fields (CRF) to model the conditional distribution of the pixels in each layout zone
- Multivariable Gaussian mixture model to learn prior-probability of the layout
- Interactive Pattern Recognition (IPR) framework to take advantage of user interaction

CRFs

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

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

CRFs



CRFs are a ng
structured

P (2)

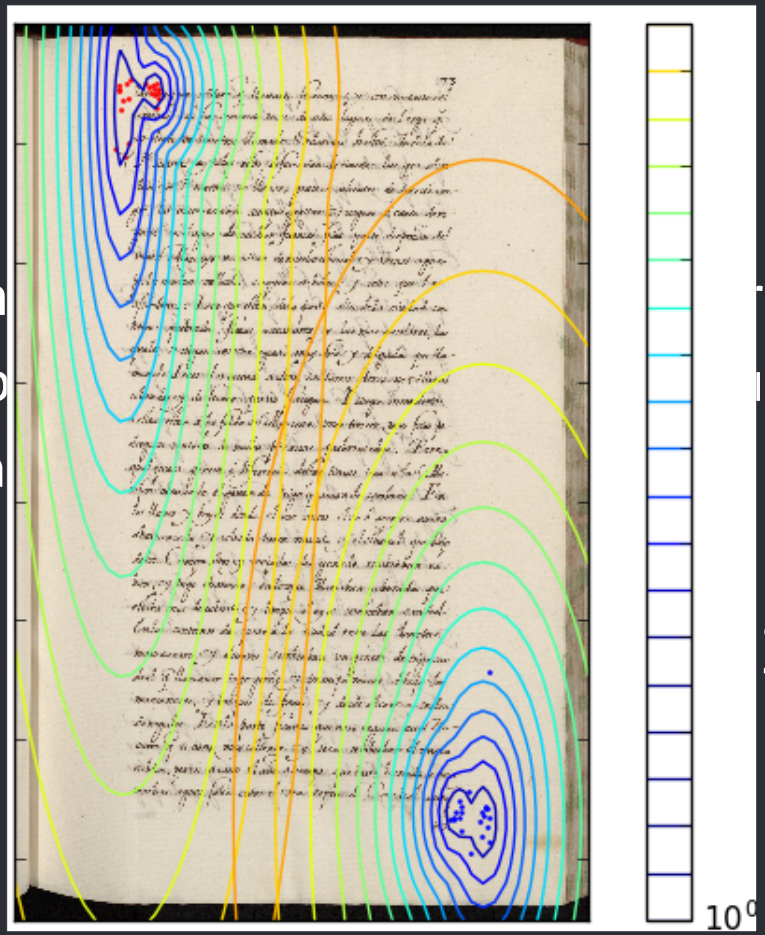
GMM

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.

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

GMM

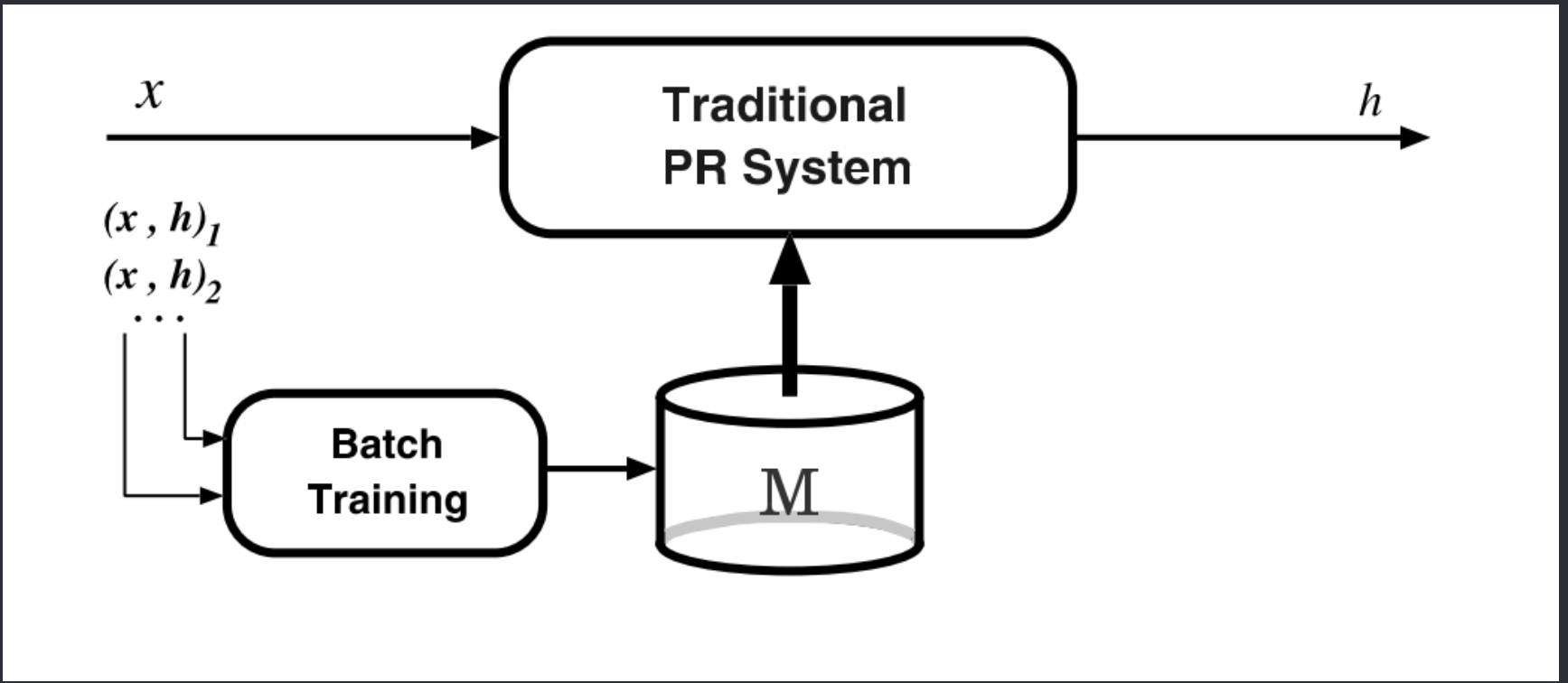
Gaussian mixture m r representing the presence of subpop lation, by means of a set of Gaussian



$$P(h) \qquad \Sigma_i) \qquad (3)$$

Classical Pattern Recognition (PR)

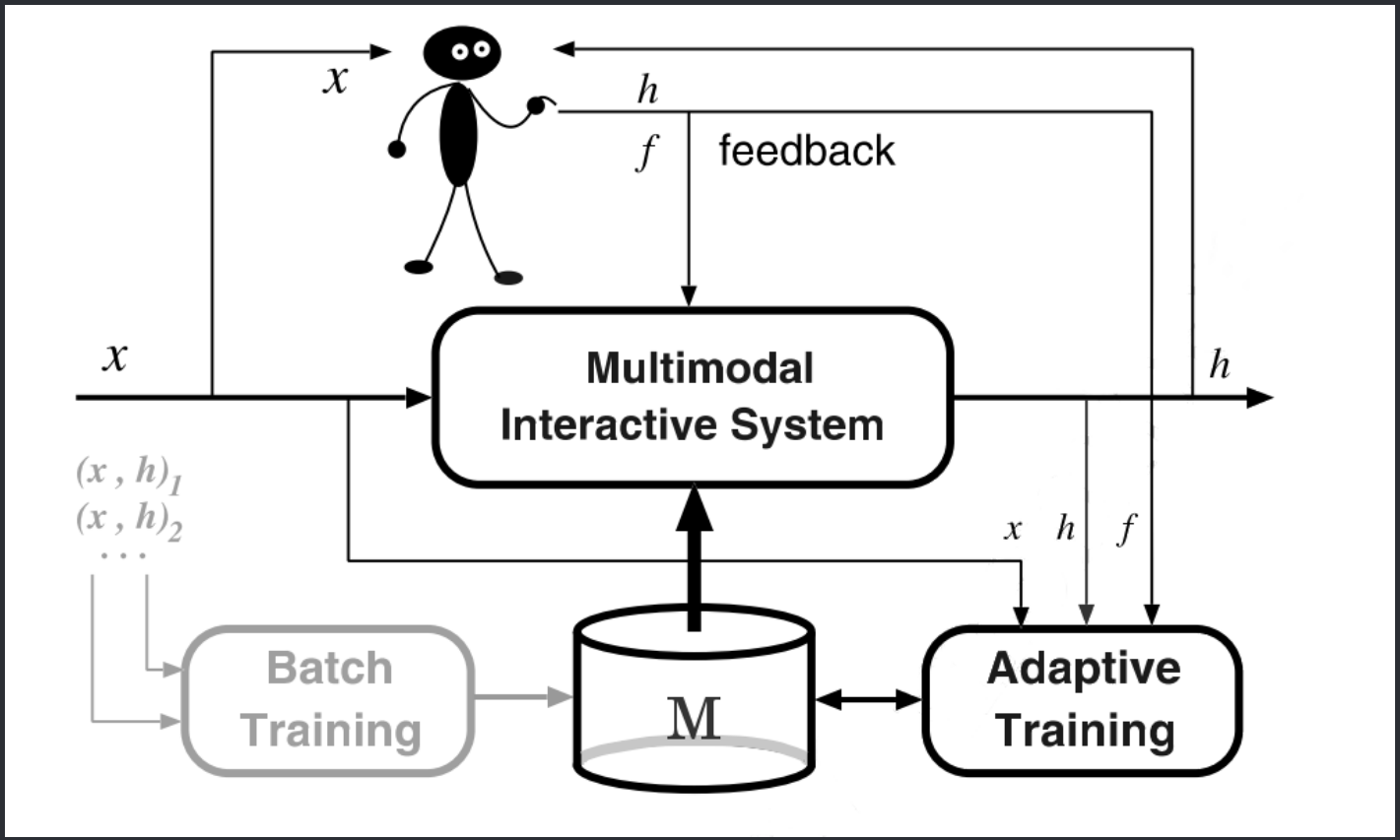
In classical PR, x is an input stimulus, observation or signal and h is a hypothesis or output, which the system derives from x.



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.



Interactive Layout Analysis

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

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

$$\Downarrow$$

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

Interactive Layout Analysis

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

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

$$\Downarrow$$

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

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

$$\log \hat{h} \approx \arg\max_{h \in \mathcal{H}} \sum_{k=1}^{K} \left(\log \underset{\mathcal{M}_g}{\mathrm{P}}(\boldsymbol{u}_k|h', f) + \log \underset{\mathcal{M}_g}{\mathrm{P}}(\boldsymbol{b}_k|h', f) + \sum_{d=1}^{D_k} \log \mathrm{P}(s_d|(\boldsymbol{u}_k, \boldsymbol{b}_k), h', f) \right) \tag{6}$$

Outline

Evaluation Measures

- Pixel-wise F1

$$P = \frac{TP}{TP + FP} \quad (7) \qquad R = \frac{TP}{TP + FN} \quad (8) \qquad F1 = \frac{2 \cdot P \cdot R}{P + R} \quad (9)$$

- MatchScore

$$\text{MatchScore}(i, k) = \frac{\text{T}(h_i^* \cap h_k)}{\text{T}(h_i^* \cup h_k)} \quad (10) \qquad FM_{i,k} = \frac{2\text{T}(h_i^* \cap h_k)}{\text{T}(h_i^*) + \text{T}(h_k)} \quad (11)$$

- Goal-Oriented Performance

$$I_{ik} = \begin{cases} h_i^* \cap h_k & \text{if } h_i^* \cap h_k \neq \emptyset \text{ and } \frac{\text{T}(I_{ik})}{\text{T}(h_k)} > T_a \\ \emptyset & \text{otherwise} \end{cases} \quad (12) \qquad SR = \frac{\sum_{i=1}^{|h^*|} \sum_{k=1}^{K} w_{ik} \times \text{T}(I_{ik})}{\sum_{i=1}^{|h^*|} \text{T}(h_i^*)} \quad (13)$$

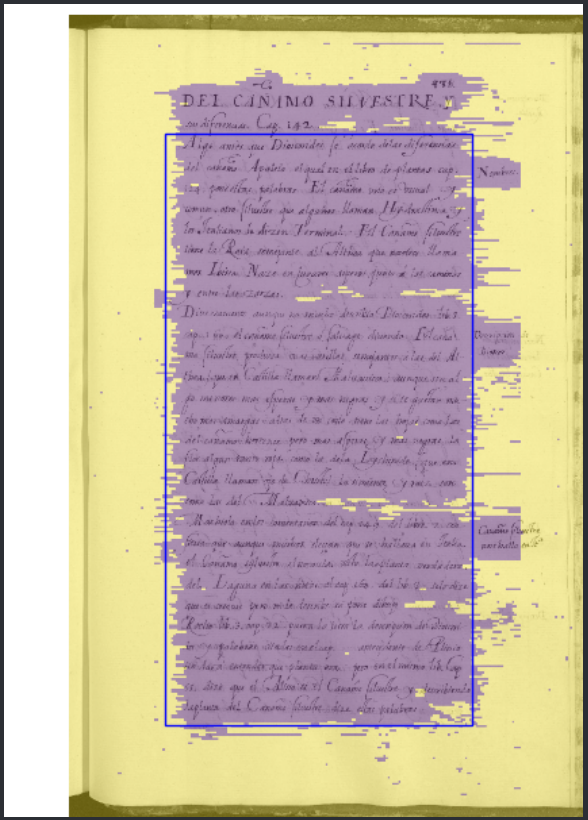
Corpus Overview

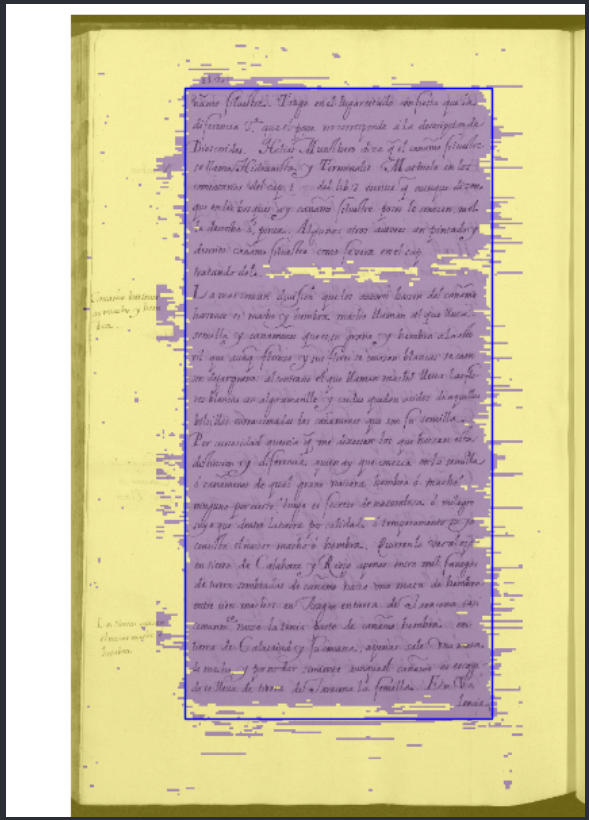
- Manuscript entitled "Historia de las Plantas" by Bernardo Cienfuegos (XVII century)
- First volume has 1 035 pages, along with their respective ground-truth layout in PAGE XML format
- Divided in seven categories, namely: catch-word, heading, marginalia, page-number, paragraph, signature-mark, and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

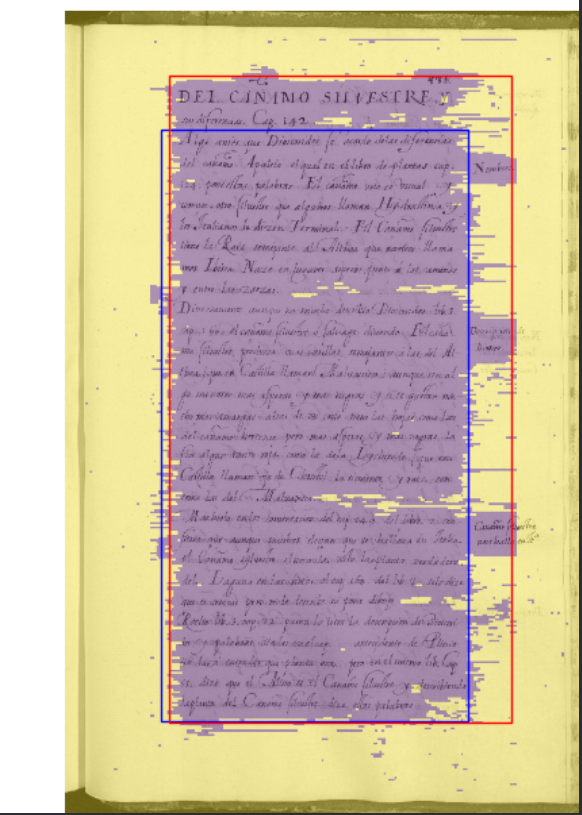


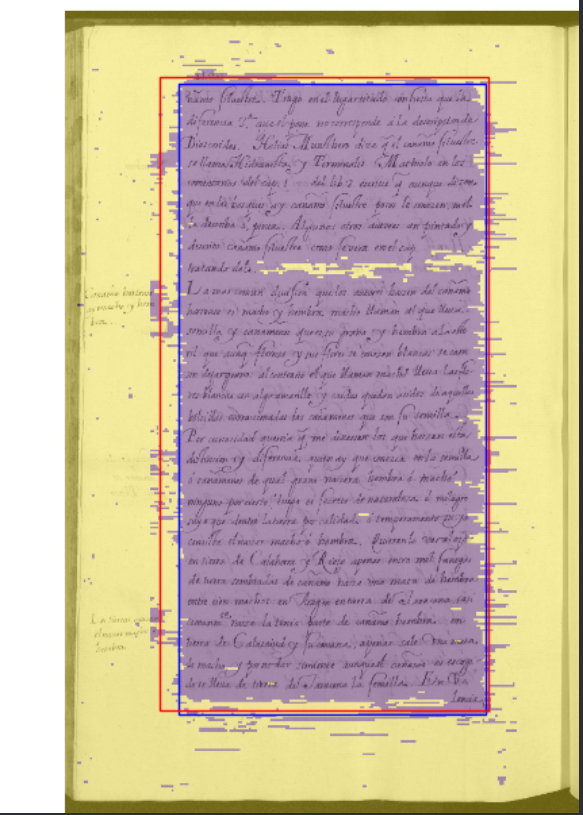


Average Results

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

Connected Components Labeling Results



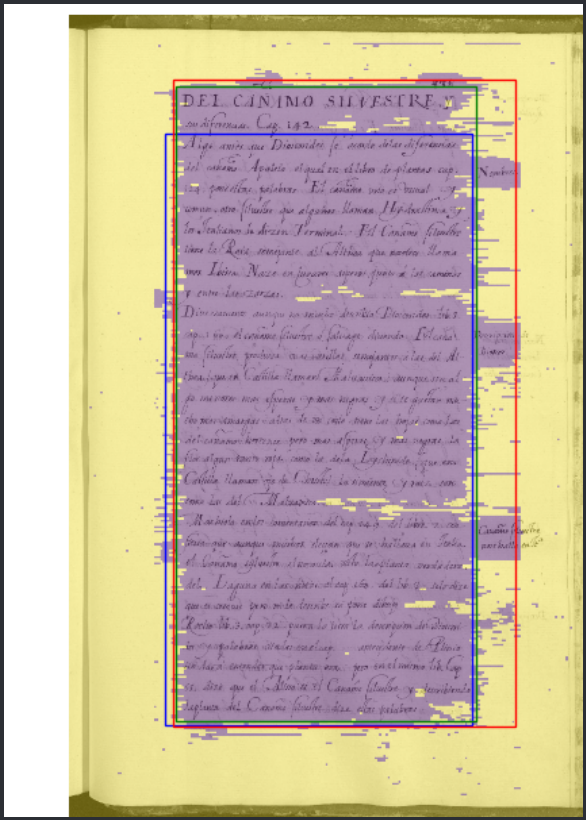


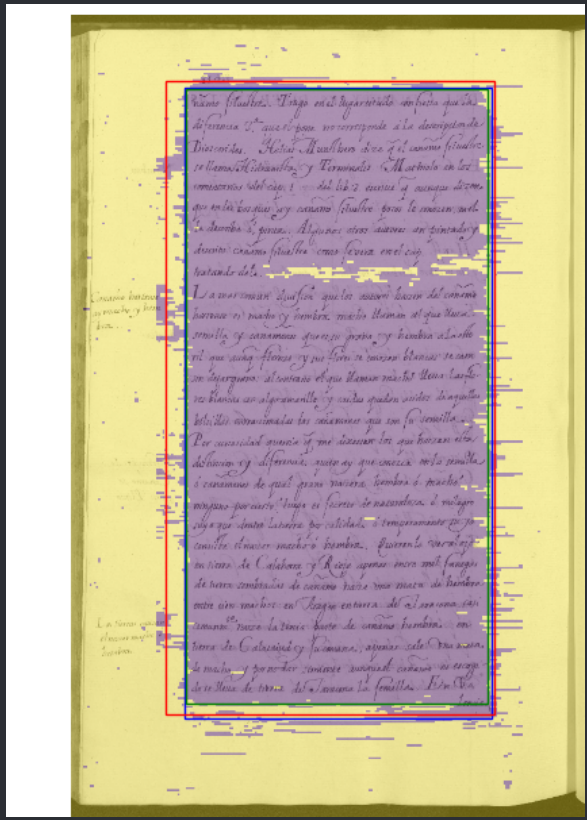
Average Results

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

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.

[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.

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.

[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.

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.

[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.

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.

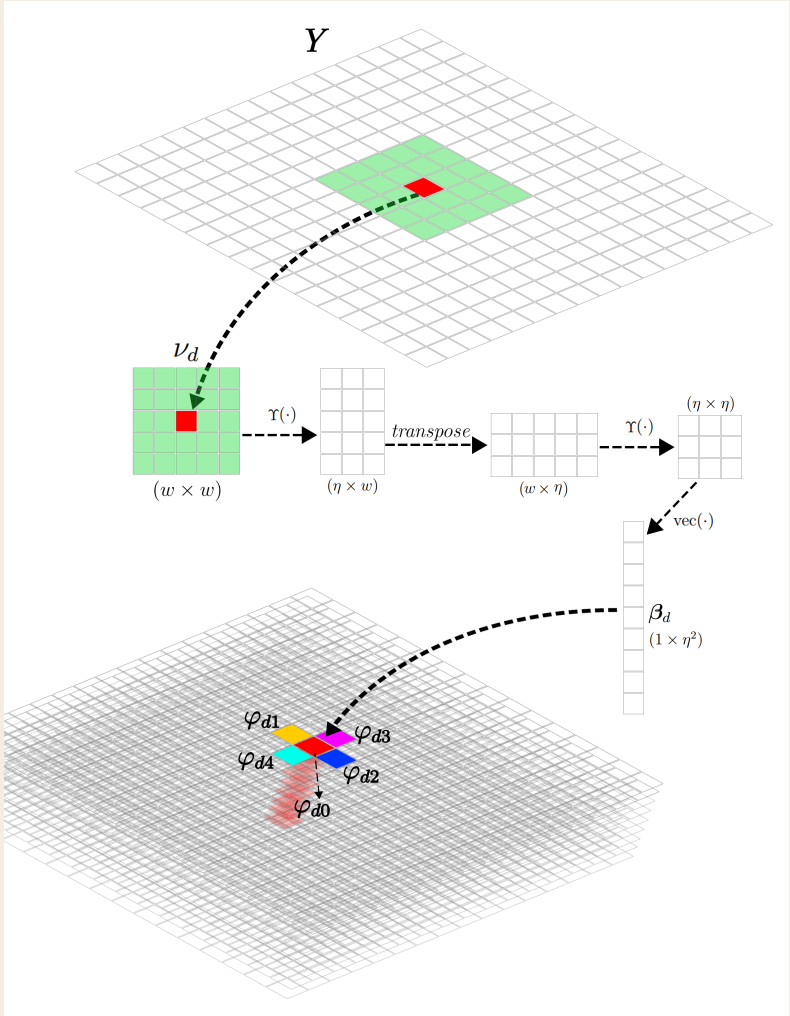
[8] A. M. Vil'kin, I. V. Safonov, and M. A. Egorova. "Algorithm for Segmentation of Documents Based on Texture Features". In: *Pattern Recognit. Image Anal.* 23.March, 2013 (2013), pp. 153–159. DOI: 10.1134/S1054661813010136. URL: http://dx.doi.org/10.1134/S1054661813010136.

BACKUP

CRFs Features

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

- 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

$$\sum_{d=1}^{D_k} \log \mathrm{P}(s_d|(\boldsymbol{u}_k, \boldsymbol{b}_k)) = \boldsymbol{I}_k[\boldsymbol{b}_k] + \boldsymbol{I}_k[\boldsymbol{u}_k] - \boldsymbol{I}_k[\boldsymbol{u}_{kr}, \boldsymbol{b}_{kc}] - \boldsymbol{I}_k[\boldsymbol{b}_{kr}, \boldsymbol{u}_{kc}] \tag{15}$$

Matrix operations

$$
\begin{aligned}
\hat{h}_{\boldsymbol{b}_1} &= \arg\min \boldsymbol{I}_0[n,m] \\
&\quad - (\boldsymbol{I}_0[f_r : n, f_c : m] + \boldsymbol{I}_0[f_r - 1, f_c - 1] - \boldsymbol{I}_0[f_r - 1, f_c : m] - \boldsymbol{I}_0[f_r : n, f_c - 1]) \\
&\quad + (\boldsymbol{I}_1[f_r : n, f_c : m] + \boldsymbol{I}_1[f_r - 1, f_c - 1] - \boldsymbol{I}_1[f_r - 1, f_c : m] - \boldsymbol{I}_1[f_r : n, f_c - 1]) \\
&\quad + \boldsymbol{P}_{\boldsymbol{u}_1}[f_r, f_c] + \boldsymbol{P}_{\boldsymbol{b}_1}[f_r : n, f_c : m] \\
\hat{h}_{\boldsymbol{u}_1} &= \arg\min \boldsymbol{I}_0[n,m] \\
&\quad - (\boldsymbol{I}_0[f_r, f_c] + \boldsymbol{I}_0[0 : f_r - 1, 0 : f_c - 1] - \boldsymbol{I}_0[0 : f_r - 1, f_c] - \boldsymbol{I}_0[f_r, 0 : f_c - 1]) \\
&\quad + (\boldsymbol{I}_1[f_r, f_c] + \boldsymbol{I}_1[0 : f_r - 1, 0 : f_c - 1] - \boldsymbol{I}_1[0 : f_r - 1, f_c] - \boldsymbol{I}_1[f_r, 0 : f_c - 1]) \\
&\quad + \boldsymbol{P}_{\boldsymbol{u}_1}[1 : f_r, 1 : f_c] + \boldsymbol{P}_{\boldsymbol{b}_1}[f_r, f_c]
\end{aligned} \tag{16}
$$