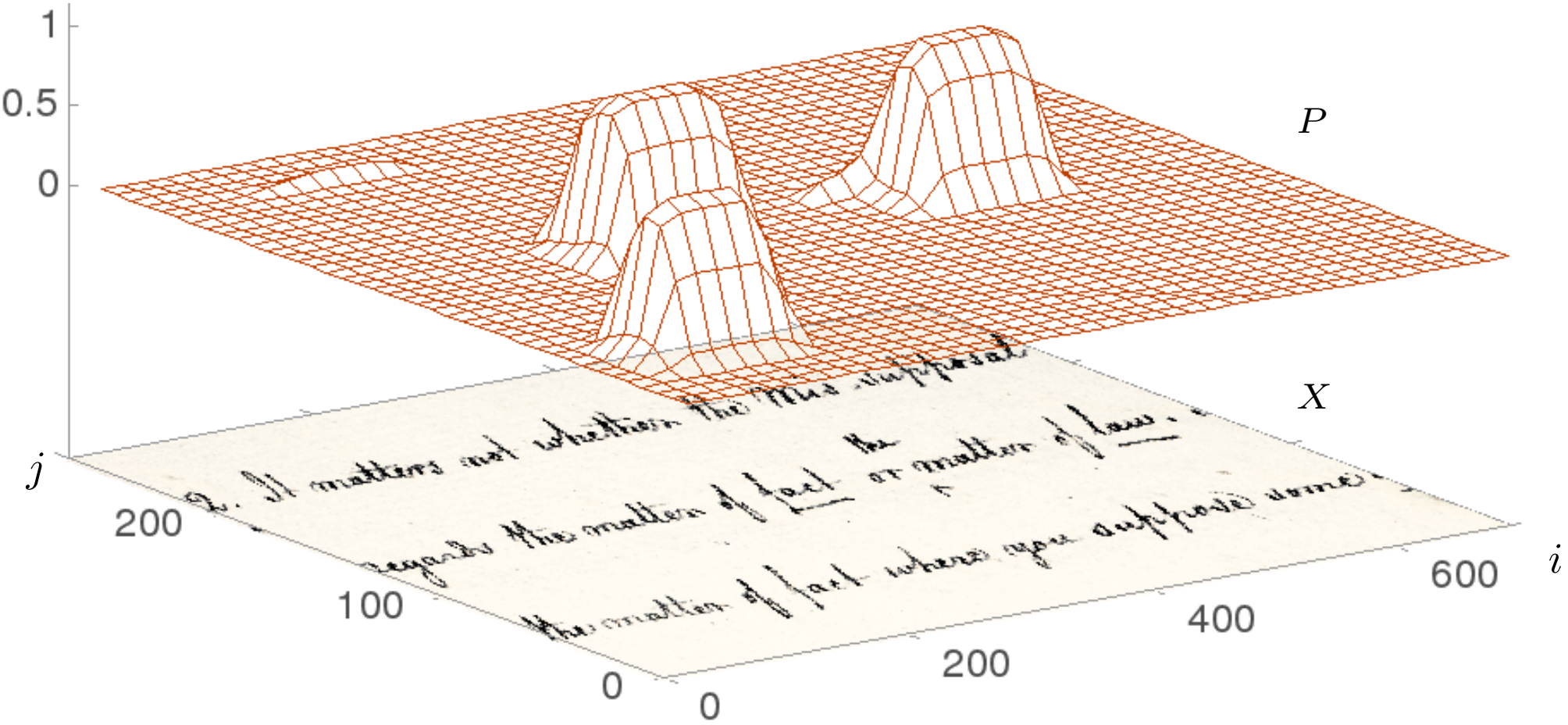
Handwritten Text Images Indexing and Search: motivation

- There are many massive text image collections out there, but their textual content remains practically inaccessible
- If perfect or sufficiently accurate text image transcripts were available, image textual content could be strightforwardly indexed for high-quality plaintext searching.
- However, manual or even interactive-predictive assited transcription is entierely prohibitive to deal with massive image collections
- Even though fully automatic transcription results are becomming increassingly good, they will probably never reach the level of accuracy needed for high-quality text indexing and search purposes
- *Good news*: indexing and search can be directly implemented on the images themselves, *without explicitly resorting to any image transcripts*, as we will see now.

Indexing and Search for Handwritten Text Images: Pixel-level Posteriorgram

Pixel-level posterior probabilities P for a text image X and word $v =$"**matter**".



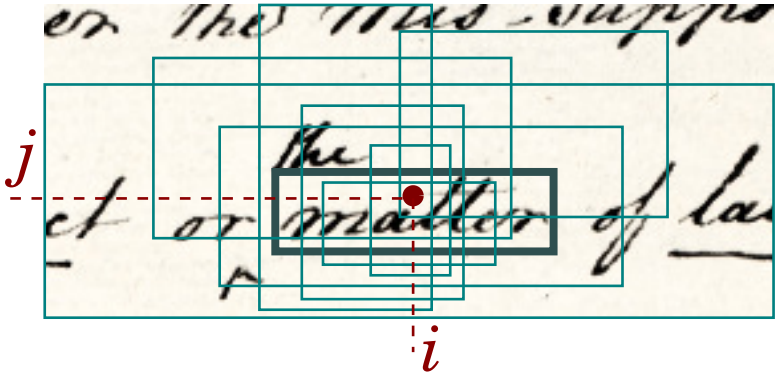
An accurate, contextual (n-gram based) *word classifier* was used to compute P. This led to high probabilities in a region of X around $(i = 25, j = 30)$, despite the rather faint appearance of "**matter**" there, but just very low probabilities in the region around $(i = 100, j = 200)$, where a very similar word, "**matters**", is written.

Computing the 2-D Posteriorgram of a given text image, X

- $P(v \mid X, i, j)$, is the probability that a word v is writen in some word-sized box B of X which includes the pixel (i, j)
- The box B is unknow, but $P(v \mid X, i, j)$ can be computed by *margizalization*:

$$P(v \mid X, i, j) \;=\; \cdots \;\propto \sum_{B \in \mathcal{B}(i,j)} P(v \mid X, B)$$

where $\mathcal{B}(i, j)$ is a set of boxes or subimages of X which include the pixel (i, j)

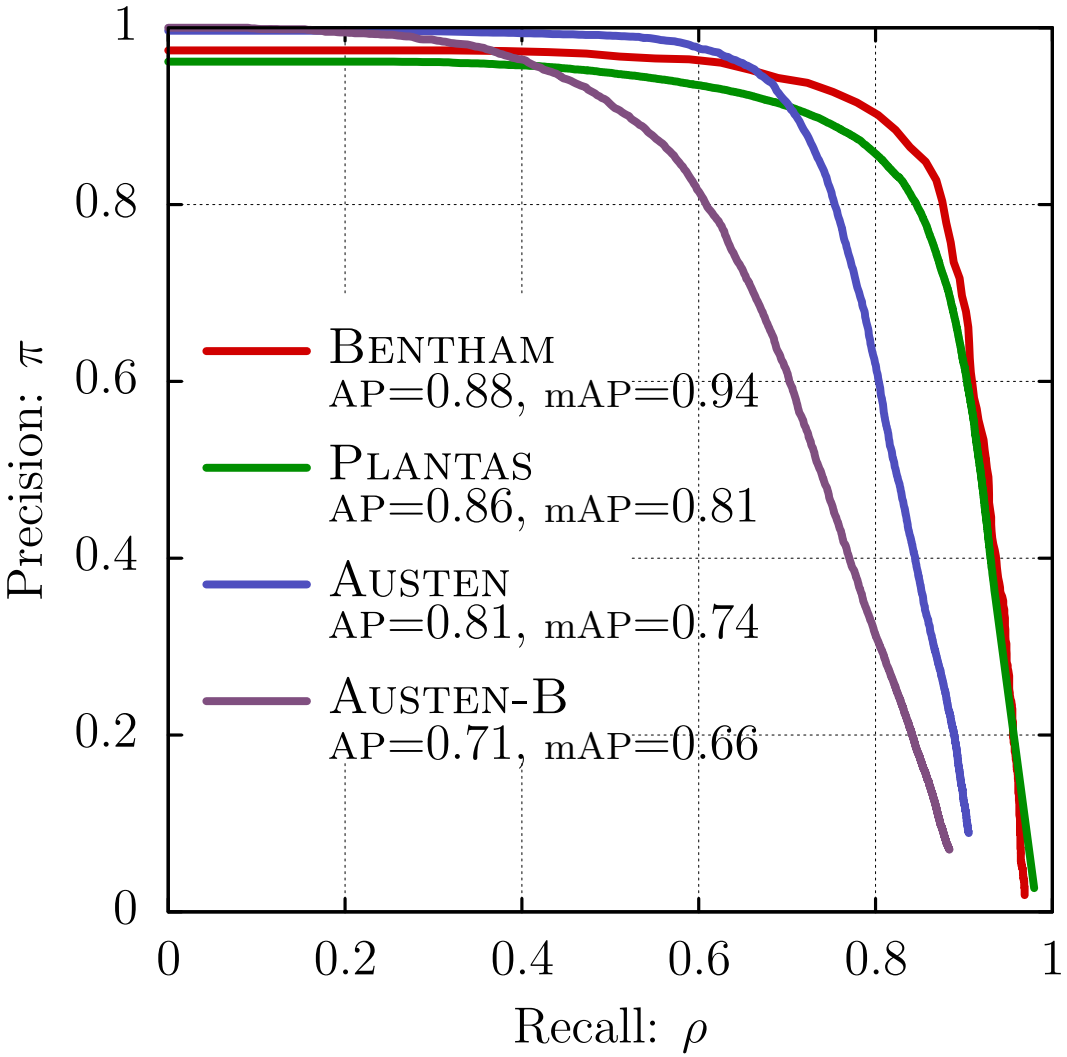


A few possible boxes $B \in \mathcal{B}(i, j)$. For $v =$ "**matter**", the thick-line box provide the highest value of $P(v \mid X, B)$, while most of the other boxes contribute only (very) low values to the sum.

- $P(v \mid X, B)$ is the posterior probability (implicitly or explicitly) computed by any *isolated word image classifier*
- *Directly computing a full 2-D posteriorgram in this way is prohibitive, but it becomes feassible by clever combinations of subsampling of (i, j) and choices of $\mathcal{B}(i, j)$.*

Laboratory Results on some tranScriptorium Data Sets

– *Recall-Precision* curves
– *Average Precision* (AP)
– *Mean Average Precision* (mAP)



Datasets training and test details

- **BENTHAM:** *Multi-hand*. *Training*: 400 pages from Bentham, 87 char. HMMs, 2-gram LM trained on Bentham texts; Lexicon 9 341 tokens. *Test*: 33 pages; query set: 6 962 keywords
- **PLANTAS (VOL-I):** *Single hand*. *Training*: 224 pages from *Plantas*, 77 char. HMMs, 2-gram LM trained with the training set + book glossary transcripts. Lexicon 11 561 tokens. *Test*: 647 pages; query set: 9 945 keywords
- **AUSTEN:** *Single hand*. *Training*: 50 Austen pages, 81 char. HMMs, 2-gram LM trained on Austen texts; Lexicon 20K tokens. *Test*: 78 pages; query set: 2 281 keywords
- **AUSTEN-B:** *Single hand*. ***No training***; using Bentham character HMMs, lexicon and LM. *Test* & query set: Same as for **AUSTEN**