SEGMENTATION-BASED HISTORICAL HANDWRITTEN WORD SPOTTING USING DOCUMENT-SPECIFIC LOCAL FEATURES

KONSTANTINOS ZAGORIS[1,2]

IOANNIS PRATIKAKIS[1]

BASILIS GATOS[2]

[1] Visual Computing Group
Democritus University of Thrace
Dept. of Electrical and Computer Engineering
Xanthi, Greece

[2] National Centre of Scientific Research "Demokritos"
Institute of Informatics and Telecommunications
Athens, Greece

WHAT IS KEY WORD SPOTTING?

- It is the task of identifying locations on a document image which have high probability to contain an instance of a queried word
- **without** explicitly recognizing it.
- It is related to Content-Based Image Retrieval systems.
- Searching a word image from a set of unindexed document images using the image content as the only information source.

CURRENT LITERATURE TRENDS

- Currently there are **two** distinct trends.

(i) Segmentation-based and **(ii) Segmentation-free** approaches.

- Their fundamental difference concerns the **search space**
  - segmented word images (segmentation-based)
  - complete document image (segmentation-free).

We address the word spotting problem with a **segmentation-based approach.**

PREVIOUS LITERATURE

Rath and Manmatha calculate two families of feature sets.

- **scalar** type features that include **aspect ratio, area, etc**.
- **profile-based features** that are based on h**orizontal and vertical words projections and the upper and lower word profiles.**

Zagoris et. al. created a similar set of profile-based features but:

- encoded **Discrete Cosine Transformation** and
- quantize through the **Gustafson - Kessel fuzzy algorithm**.

Rodriguez and Perronnin extract features from a sliding window, based on the **first gradient** and inspired by the **SIFT keypoint descriptor**.

BAG-OF-VISUAL WORDS MODEL

Recently, there was an influx of works based on the local features in the form of the Bag-of-Visual Words model.

Llados et. al. evaluate the performance of various word descriptors :

- a bag of visual words procedure (BoVW),
- a pseudo-structural representation based on Loci Features,
- a structural approach by using words as graphs, and
- sequences of column features based on DTW.

They found that the statistical approach of the BoVW produces the best results, although the memory requirements to store the descriptors are significant.

PROBLEMS WITH CURRENT LOCAL FEATURES

Most works using local features are based on the Scale Invariant Feature Transform (SIFT) in order to describe the local information

- The original application of these local features are the natural images which they have many structural differences compared to document images
- The detection of the most powerful edges through pyramid scaling creates local points between text lines.
- Invariant properties in the descriptor results in noise amplification so they are more sensitive to the noise and the complex texture of the background.

TEXTURE VS SHAPE FEATURES

Features for word spotting which rely only on word shape characteristics **are not effective** in dealing with a document collection created by different writers, containing significant writing style variations.

Although slant and skew preprocessing techniques can reduce the shape variations, they cannot eliminate the problem as the whole structure of the word is different in most of the cases.

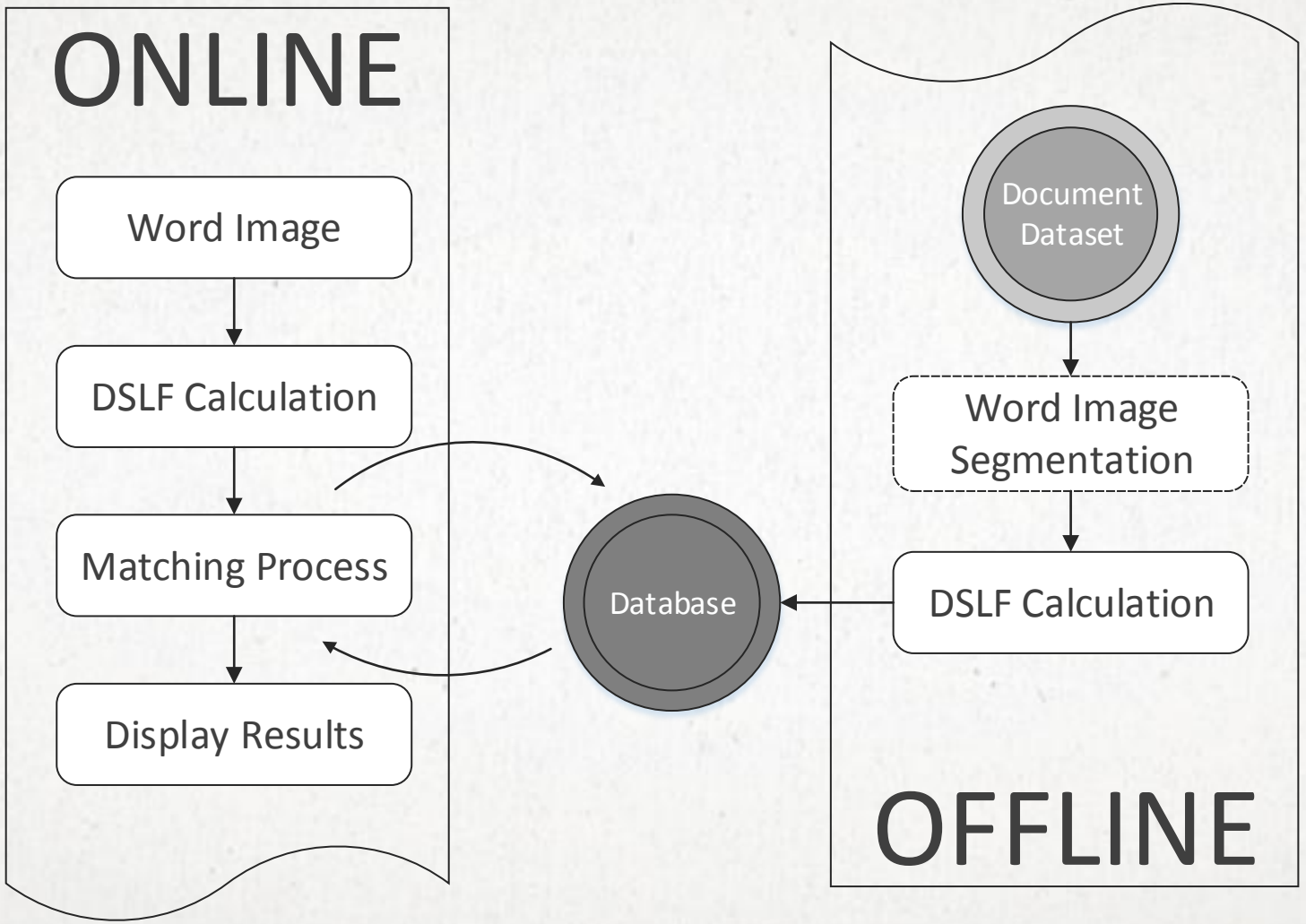
In this respect, we argue that although the shape information is meaningful, the **texture information in a spatial context** is more reliable.

DOCUMENT SPECIFIC LOCAL FEATURES (DSLF)

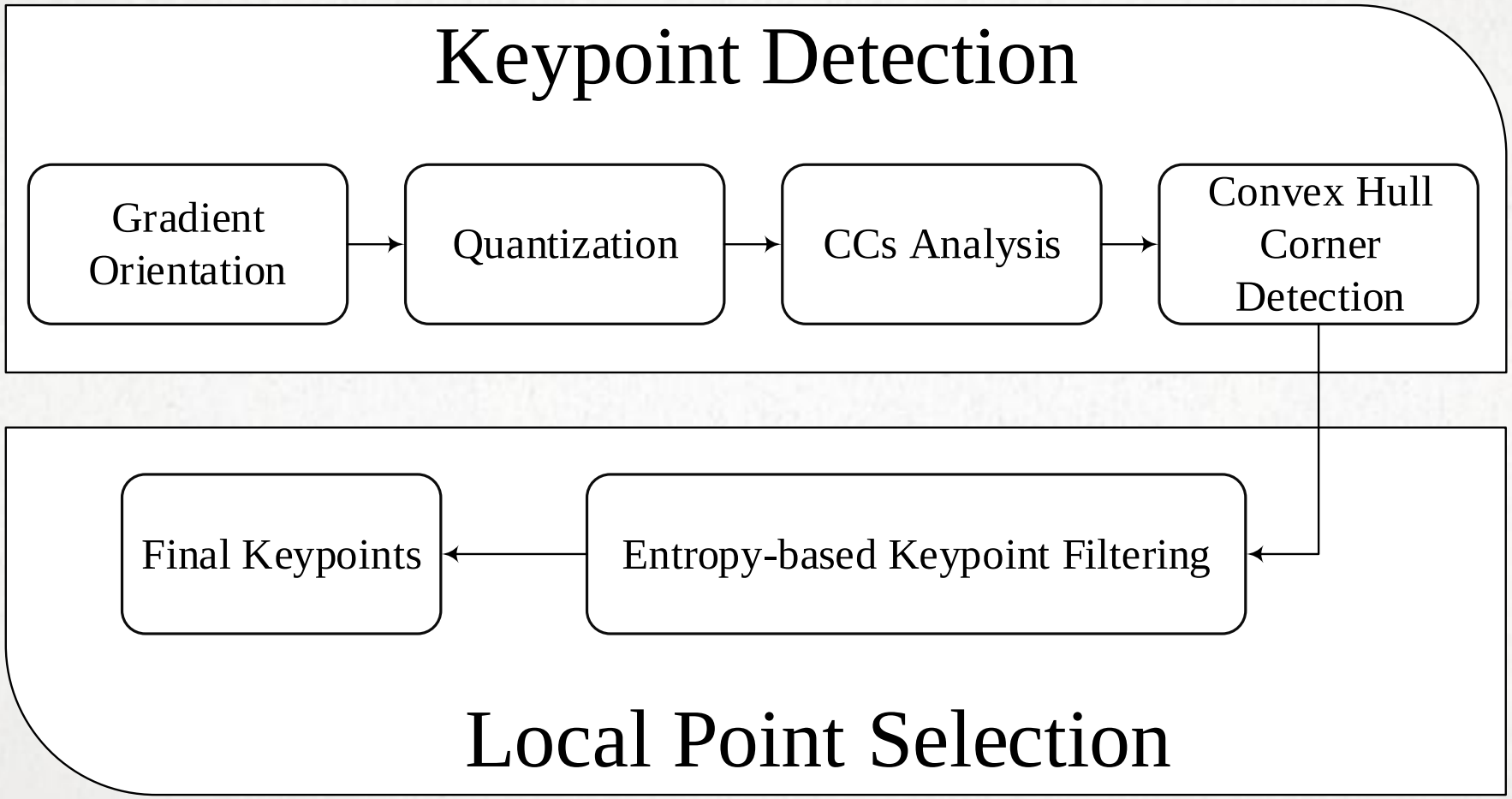
Taking into account the aforementioned considerations, we propose:

- novel local features which are specific for documents and a
- matching procedure that does not rely on codebook creation (as on BoVW).

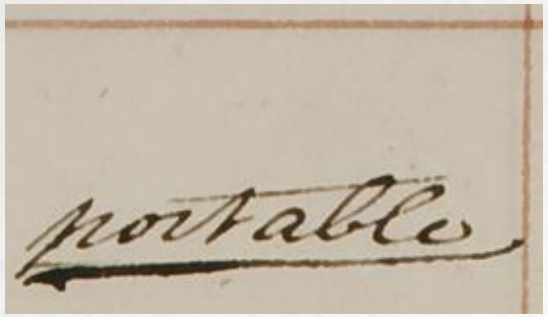
PROPOSED WORD SPOTTING FRAMEWORK



KEYPOINT DETECTION AND SELECTION

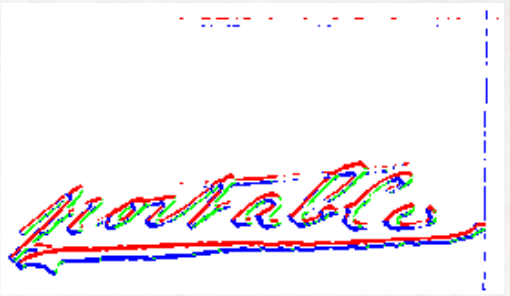


KEYPOINT DETECTION AND SELECTION

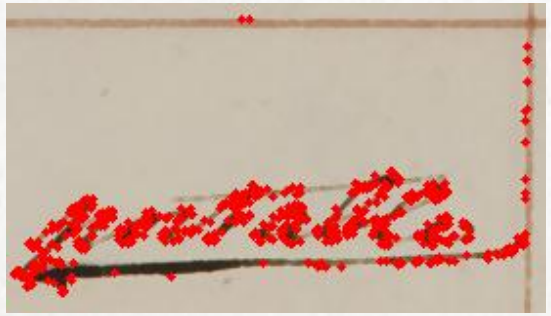


original document image

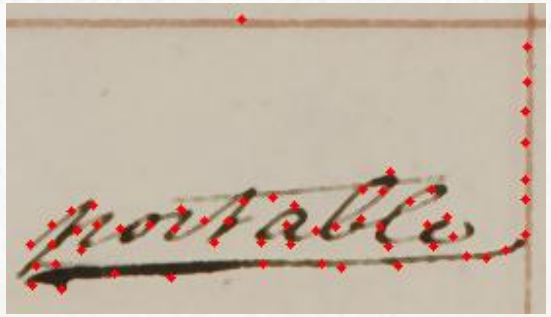
orientation of the gradient vector



quantization of the gradient vector orientation



initial keypoints



final keypoints

FEATURE EXTRACTION

- The feature for the local keypoint is calculated upon the quantized gradient angles
- An area of 18x18 pixels around the kP, is divided into 9 cells with size 6x6 for each of them.
- Each cell is represented by a 3-bin histogram (each bin corresponds to a quantization level).
- Each pixel accumulates a vote in the corresponding angle histogram bin. The strength of voting depends on the norm of the gradient vector and on the distance from the location of local point as shown at the following equation:

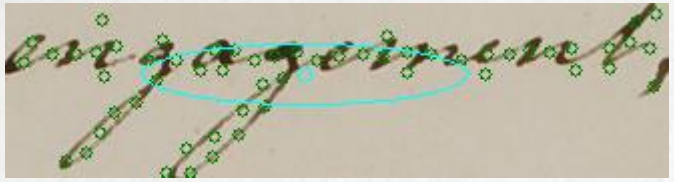
$$V_{x,y} = s_{x,y} \cdot \|G_{x,y}\| \qquad s_{x,y} = 1 - \frac{2}{3} \cdot \frac{\sqrt{(x - x_{LP})^2 + (y - y_{LP})^2}}{9\sqrt{2}}$$

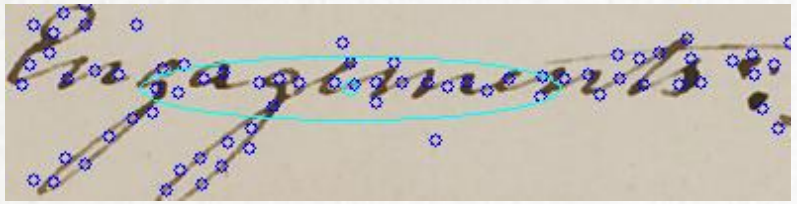
- The task of the $s_{x,y}$ variable is to weigh the pixel participation to the histogram taking into account its distance from the kP.

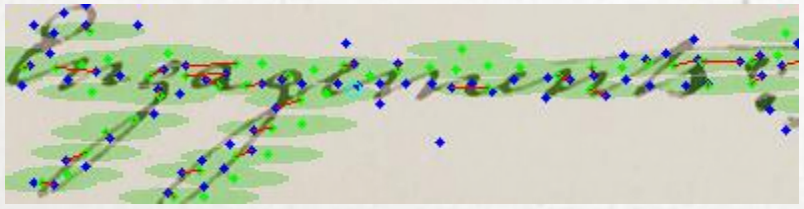
MATCHING PROCEDURE

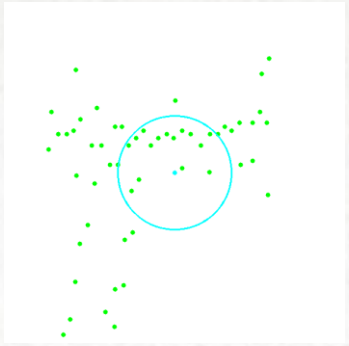
- In the case of segmentation-based word spotting, the aim is to match the query keypoints to the corresponding keypoints of any word image in the document.
- Local Proximity Nearest Neighbor (LPNN) search is implemented.
- The advantage of LPNN search is two-fold:
  - it enables a search in focused areas instead of searching in a brute force manner and
  - it goes beyond the typical use of a descriptor by the incorporation of spatial context in the local search addressed.

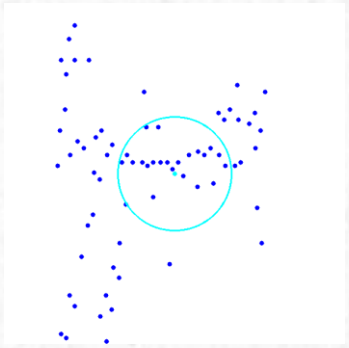
MATCHING PROCEDURE

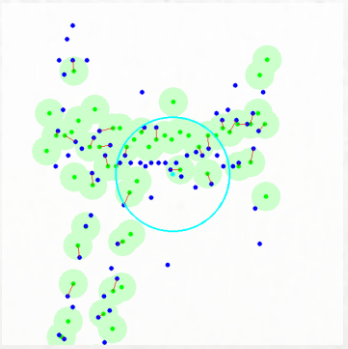












Update the location for each keypoint to a new normalized space:

$$p_x^{i\prime} = \frac{p_x^i - c_x}{D_x}, p_y^{i\prime} = \frac{p_y^i - c_y}{D_y}$$

where:

$$(c_x, c_y) = \left(\frac{\sum_{i=1}^{k} p_x^i}{k}, \frac{\sum_{i=1}^{k} p_y^i}{k}\right)$$

$$D_x = \frac{\sum_{i=1}^{k} |p_x^i - c_x|}{k}, D_y = \frac{\sum_{i=1}^{k} |p_y^i - c_y|}{k}$$

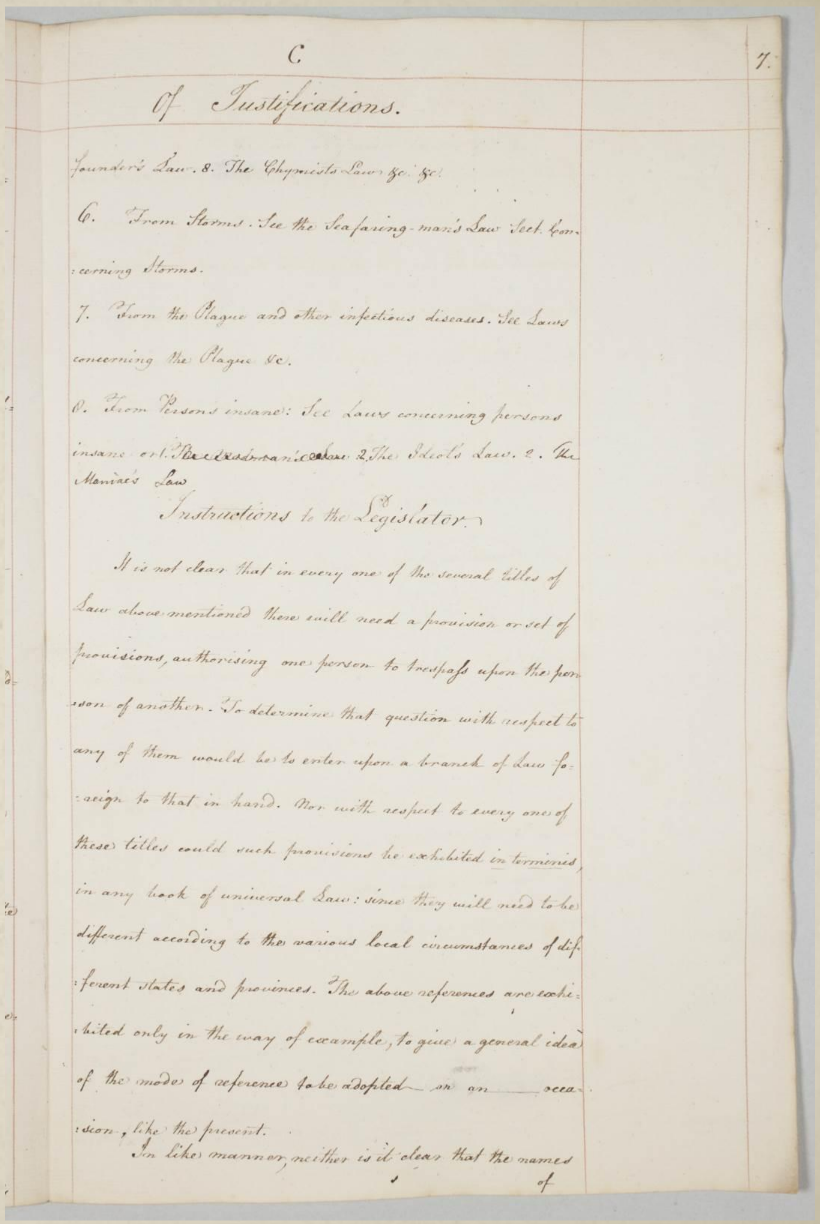
k denotes the total number of the keypoints in a word image.

EVALUATION - DATASETS

BENTHAM DATASET

- It consists of **50** high quality (approximately 3000 pixel width and 4000 pixel height) handwritten manuscripts written by Jeremy Bentham (1748-1832).

- The variation of the same word is extreme and involves writing style, font size, noise as well as their combination.



C 7.

Of Justifications.

Founder's Law. 8. The Chymists Law &c. &c.

6. From Storms. See the Seafaring-man's Law Sect. Con-
:cerning Storms.

7. From the Plague and other infectious diseases. See Laws
concerning the Plague &c.

8. From Persons insane: See Laws concerning persons
insane — 1. [illegible] 2 The Ideot's Law. 2. The
Maniac's Law

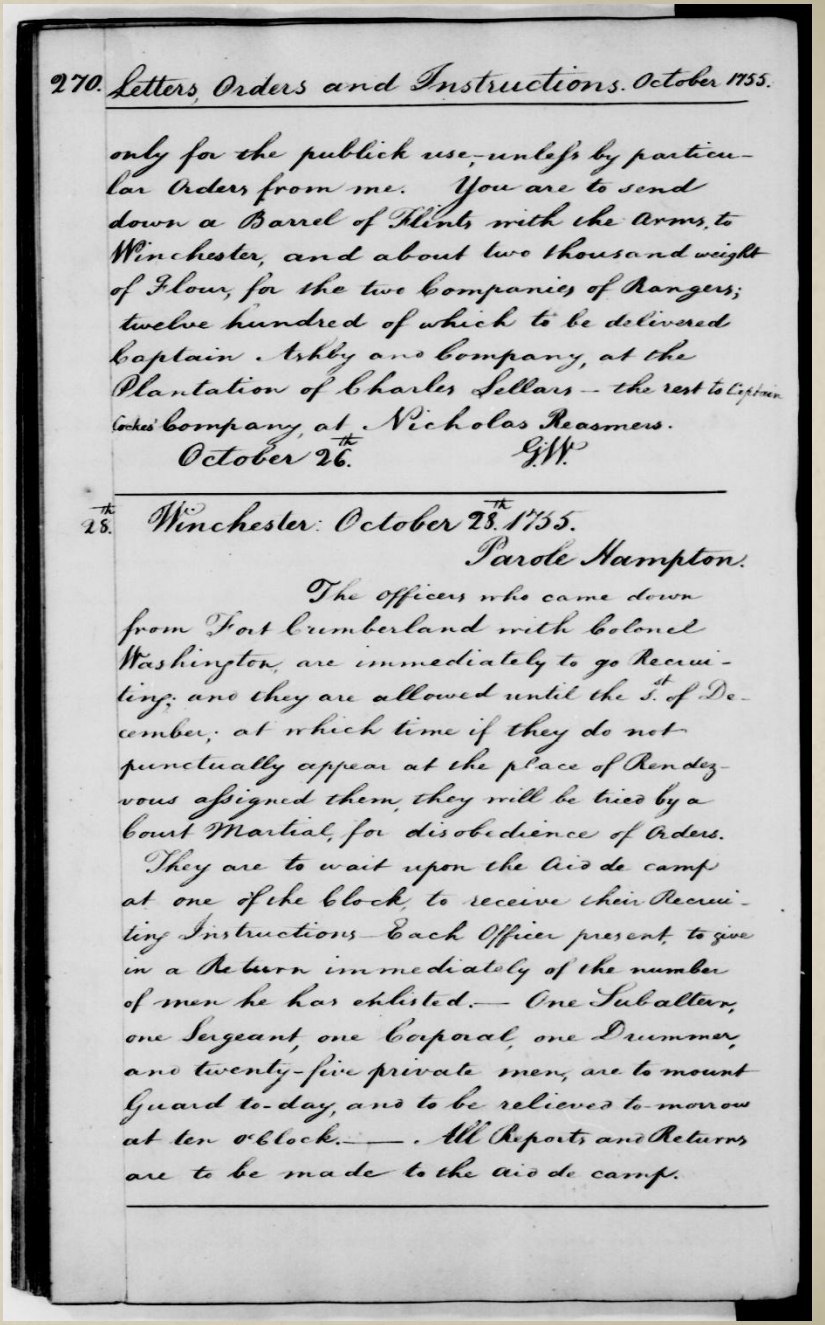
Instructions to the Legislator.

It is not clear that in every one of the several titles of
Law above-mentioned there will need a provision or set of
provisions, authorising one person to trespass upon the per-
son of another. To determine that question with respect to
any of them would be to enter upon a branch of Law fo-
:reign to that in hand. Nor with respect to every one of
these titles could such provisions be exhibited in terminis
in any book of universal Law: since they will need to be
different according to the various local circumstances of dif-
:ferent states and provinces. The above references are exhi-
:bited only in the way of example, to give a general idea
of the mode of reference to be adopted on an occa-
:sion, like the present.
In like manner, neither is it clear that the names
of

EVALUATION - DATASETS

WASHINGTON DATASET

- It consists of **20** document images from George Washington Collection of the Library of Congress
- The documents are were scanned from microfilm in 300 dpi resolution.



270. Letters, Orders and Instructions. October 1755.

only for the publick use, unless by particular Orders from me. You are to send down a Barrel of Flints with the Arms, to Winchester, and about two thousand weight of Flour, for the two Companies of Rangers; twelve hundred of which to be delivered Captain Ashby and Company, at the Plantation of Charles Sellars — the rest to Captain Cockes Company, at Nicholas Reasmers.

October 26th. GW.

28th Winchester: October 28th 1755.

Parole Hampton.

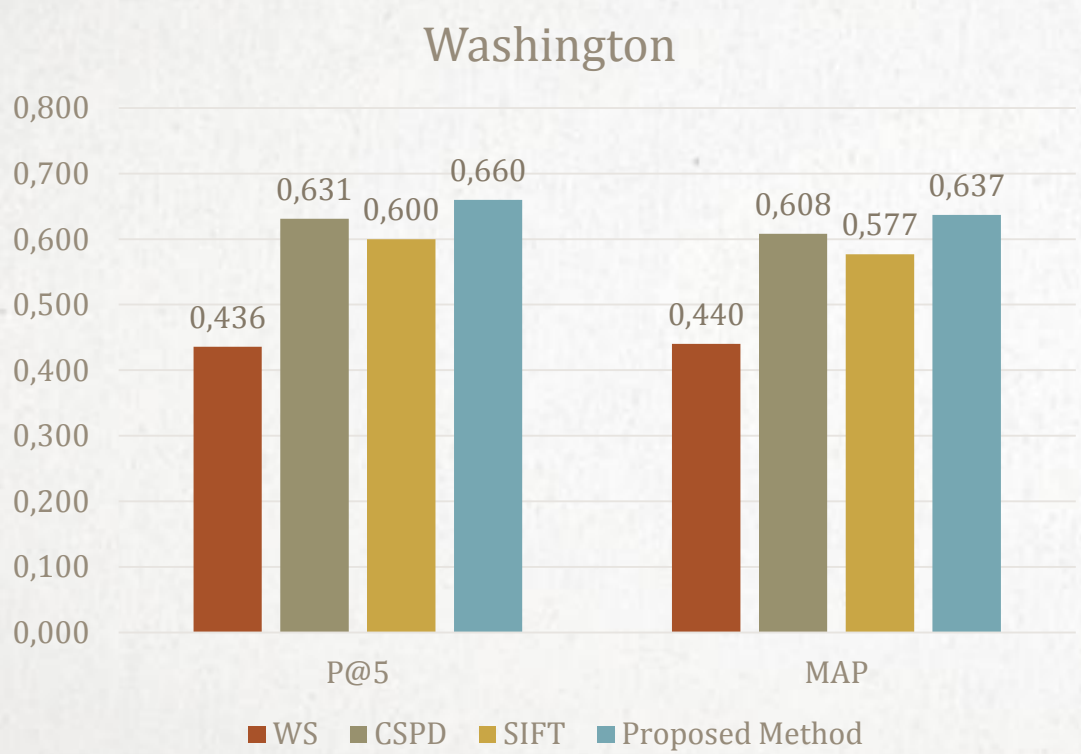
The Officers who came down from Fort Cumberland with Colonel Washington, are immediately to go Recruiting: and they are allowed until the 1st of December; at which time if they do not punctually appear at the place of Rendezvous assigned them, they will be tried by a Court Martial, for disobedience of Orders.

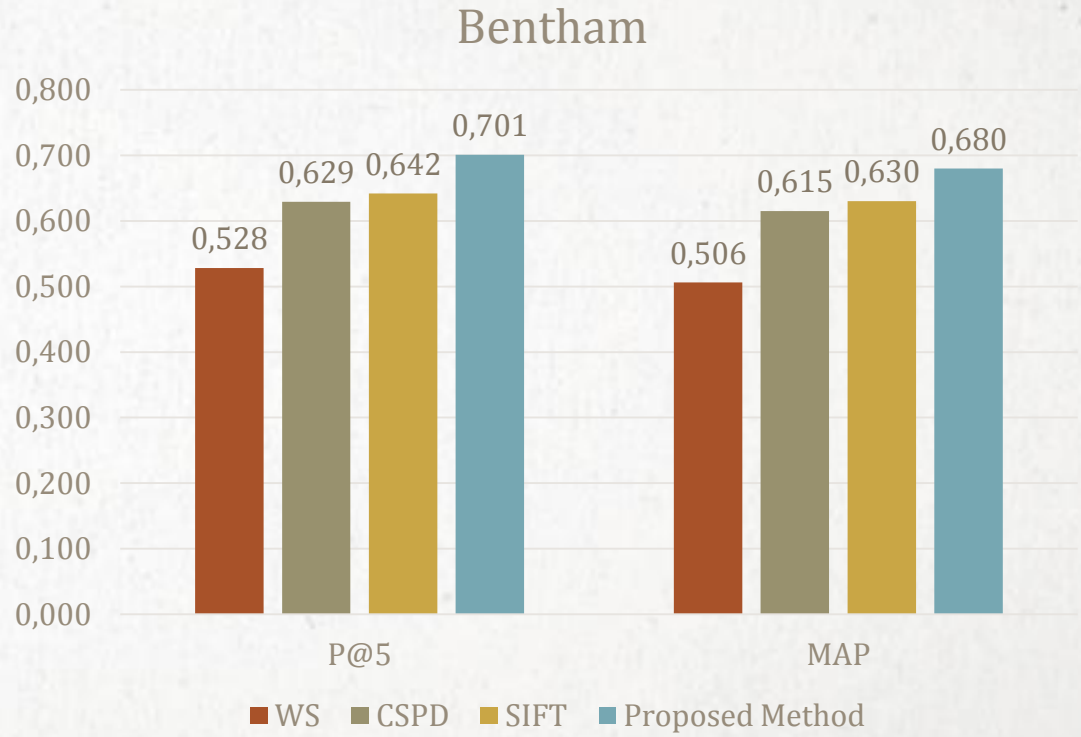
They are to wait upon the Aid de camp at one of the Clock, to receive their Recruiting Instructions — Each Officer present, to give in a Return immediately of the number of men he has enlisted. — One Subaltern, one Sergeant, one Corporal, one Drummer, and twenty-five private men, are to mount Guard to-day, and to be relieved to-morrow at ten o'Clock. — All Reports and Returns are to be made to the Aid de camp.

EVALUATION STRATEGY

- Two evaluation metrics: **Precision at the k Top Retrieved words (P@k)** and the **Mean Average Precision (MAP).**
- **P@5 is the precision at top 5 retrieved words.** This metric defines how successfully the algorithms produce relevant results to the first 5 positions of the ranking list
- **MAP** is a typical measure for the performance of information retrieval systems
- For the experiments, the word image segmentation information is taken from the ground truth corpora.
- The total word image queries for the Washington dataset was **1570** and for the Bentham dataset was **3668**.
- Both query sets contain words appearing in various frequencies and sizes
- Evaluated against two previous segmentation-based **profile-based** strategies
- Then, in order to highlight the advantage of the proposed DSLF, it was **replaced by the SIFT** but the proposed matching algorithm remained the same.

OVERALL PERFORMANCE EVALUATION RESULTS





CONCLUSION

In this work, novel local features are proposed driven by the challenges presented in historical handwritten word spotting scenarios.

The proposed method outperformed both the profile-based strategies and the SIFT local features.

Moreover, a matching procedure was presented based on Local Proximity Nearest Neighbour, that augments performance in terms of effectiveness and efficiency incorporating spatial context.

The proposed framework achieves better performance after a consistent evaluation against two profile-based approaches as well as the proposed approach with the popular SIFT local features in two different handwritten datasets.

ΕΥΧΑΡΙΣΤΩ ΠΟΛΥ!
THANK YOU!