Unsupervised Word Spotting in Historical Handwritten Document Images Using Document-Oriented Local Features

Konstantinos Zagoris, *Member, IEEE*, Ioannis Pratikakis, *Senior Member, IEEE*, and Basilis Gatos

***Abstract*— Word spotting strategies employed in historical handwritten documents face many challenges due to variation in the writing style and intense degradation. In this paper, a new method that permits effective word spotting in handwritten documents is presented that it relies upon document-oriented local features, which take into account information around representative keypoints as well a matching process that incorporates spatial context in a local proximity search without using any training data. Experimental results on four historical handwritten data sets for two different scenarios (segmentation-based and segmentation-free) using standard evaluation measures show the improved performance achieved by the proposed methodology.**

***Index Terms*— Word spotting, handwritten documents, local features.**

I. Introduction

DIGITIZED historical manuscripts are not fully exploited due to lack of proper browsing and indexing tools. Traditional approaches in document indexing involving Optical Character Recognition (OCR) which although performs well in modern printed documents does not operate effectively in the case of historical handwritten documents which entail different document degradations due to low paper quality, writing style variations, bleeding through ink, shadows, non-uniform illumination, smears, etc.

A promising strategy to deal with unindexed documents is a word matching procedure that relies upon a low-level pattern matching called word spotting [1]. It is directly related to Content-Based Image Retrieval, since it searches a word in a set of unindexed documents using the image content as the only information source. As final outcome, the system returns to the user a ranked list of document word images based upon the similarity with the query word image. Finally, word spotting can be conceived as the task of identifying locations on a document image which have high probability to correspond to an instance of a queried word-image, without explicitly recognizing it.

In the literature, word spotting appears under two distinct strategies wherein the fundamental difference concerns the search space which could be either a set of segmented word images (segmentation-based approach) or the complete document image (segmentation-free approach). The selection of the segmentation-based strategy is preferred when the layout is simple enough to correctly segment the words while the segmentation-free strategy performs better when there is considerable degradation on the document. Nevertheless both strategies use an operational pipeline where features extraction and matching have prominent role. Although there is an abundance of systems suitable for both modern [2], [3] and historical machine-printed documents [4]–[7], very few of them are suitable for handwritten documents due to heavy degradation, variation in the writing style and text layout complexity. The challenging nature of word spotting in handwritten documents has motivated the organization of two dedicated international competitions, namely the ICFHR 2014 Handwritten Keyword Spotting Competition [8] and the ICDAR 2015 Competition on Keyword Spotting for Handwritten Documents [9] where seven (7) group totally have been competed in two different word spotting scenarios (segmentation-free and segmentation-based).

In this paper, a new method that permits effective word spotting in handwritten documents is presented that it relies upon document-oriented local features which take into account information around representative keypoints as well a matching process that incorporates spatial context in a local proximity search without using any training data.

The remainder of the paper is organized as follows: At Section II, a comprehensive review of the related work is given, Section III presents the proposed word spotting architecture,, wherein, Section III-B details the keypoints detection and features extraction. In Section IV, the matching procedure is presented. Finally, in Section V the experimental results are discussed while in Section VI conclusions are drawn.

II. Related Work

Previous approaches of word spotting can be distinguished to supervised or unsupervised based on the need of training

Manuscript received June 17, 2016; revised April 13, 2017; accepted April 29, 2017. Date of publication May 3, 2017; date of current version June 23, 2017. This work was supported in part by the EU Horizon 2020 Research and Innovation Programme 'READ' under Grant 674943. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Christos Bouganis. *(Corresponding author: Konstantinos Zagoris.)*

K. Zagoris is with the National Centre for Scientific Research "Demokritos", Institute of Informatics and Telecommunications, Athens, Greece, and also with the Democritus University of Thrace, Xanthi, 67100 Greece (e-mail: kzagoris@ee.duth.gr).

I. Pratikakis is with the Department of Electrical and Computer Engineering, Democritus University of Thrace, Xanthi, 67100 Greece (e-mail: ipratika@ee.duth.gr).

B. Gatos is with the National Centre for Scientific Research "Demokritos", Institute of Informatics and Telecommunications, 15310 Athens, Greece (e-mail: bgat@iit.demokritos.gr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2017.2700721

data or not. Furthermore, each approach is characterized by the context wherein it operates in relation to the use of word image segmentation (segmentation-based) or the lack of any dependency on segmentation (segmentation-free).

Initial efforts in unsupervised segmentation-based word spotting followed a methodological pipeline using as a first step, pre-processing including binarization and text layout analysis followed by word image segmentation. Then, analysing the segmented word image, a set of features is extracted. Based on these features, a distance is used to measure the similarity between the query word image and each of the segmented word image found in a document image or a collection of images.

In this spirit, Rath and Manmatha [10], [11] and Lavrenko *et al.* [12] calculate two families of feature sets. On the one hand, they use scalar type features that include aspect ratio, area, etc. On the other hand, the profile-based features are used that are based on horizontal and vertical words projections and the upper and lower word profiles. Dynamic Time Warping (DTW) is used in order to match different length-variant features sets.

Zagoris *et al.* [13] (CSPD) created a similar set of profile-based features, differently encoded by Discrete Cosine Transformation, normalized by the first coefficient and quantized by the Gustafson and Kessel [14] fuzzy algorithm. The result was a very short descriptor, which has been tested on a Greek handwriting database from different writers, the Washington words database and the MPEG-7 CE1 Set B database. It uses a weighted Minkowski L_1 distance to match the different features set. It has been used as the baseline approach for the Track I, Assignment-I.A. (Segmentation-based) from the ICDAR 2015 competition [9] as it is freely available.[1]

Finally, Srihari *et al.* [15] present a system for searching handwritten Arabic documents based on a set of binary shape features suitable for Arabic script along with a correlation distance that performed best for matching.

Recent works on unsupervised word spotting do not require segmentation at any level, thus denoted as segmentation-free. Leydier *et al.* [16] detects zones on the images that represent the most informative parts based on the gradient orientation calculated by the convolution of the image with the first and second derivatives of the Gaussian kernel. Their matching procedure is based on naive elastic matching. In order to create guides for the matching, they use morphological operations; an opening with a heuristic structuring element based on the document type. The evaluation experiments were conducted on medieval manuscripts of Latin and Semitic alphabets. They found that their algorithm is not suitable for short words (less than four letters) as the shorter the word, the less information to compare with. In [17], they introduced an improved version of their initial segmentation-free approach. The main revised parts comprise the matching engine and guide detection which is based on second derivatives along the isophote curves. Although it is an improvement compared to [16], the way they extract the features and match them, they are very sensitive to different writing variations and matching word with different font sizes. Moreover, their matching engine is time consuming making its use impractical for large datasets.

Gatos and Pratikakis [18] presented a block-based document image descriptor that is used in a template matching scheme. They created versions of the query image that are scaled and rotated to produce different word instances. For each word instance they calculated different set of feature vectors. Their various versions of queries creates a lot of noise in the final merging state from the different queries especially where there is variations in the writing and size of the matching words.

Kovalchuk *et al* [19] preprocessed the documents by a simple binarization process and resize them to fit a fixed-size rectangle. Then, extract from them HOG and LBP descriptors. The retrieval is performed by nearest-neighbour search, followed by a simple oppression of extra overlapping candidates. This algorithm awarded the 1^{st} prize in the Track II (Segmentation-Free) of ICFHR 2014 Handwritten KeyWord Spotting Competition [8].

Apart the aforementioned methods which operate in an unsupervised context. there exist methods that are employed in a supervised context using training data to learn similarities.

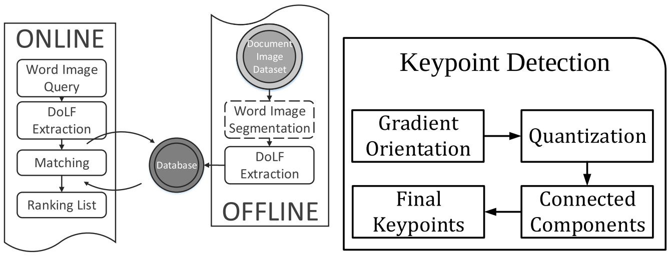
Almazan *et al.* [20] presented a method which both word images and text strings are embedded in a common vectorial subspace. In this subspace, images and strings that represent the same word are close together, allowing one to cast recognition and retrieval tasks as a nearest neighbour problem. This algorithm awarded the 1^{st} prize in the Track I (Segmentation-Based) of ICFHR 2014 Handwritten KeyWord Spotting Competition [8].

Some works operate at text line level using HMM models to spot the words [21]. Another approaches embed a discriminative stage in HMM model such as SVM [22], a neural network [23] or deep learning network architectures [24]. Although, they provide better results, they have a lot of drawbacks. Firstly, they required a lot of training data and some cannot be able to identify words that are not present in the training set. Moreover, their requirement for line or word segmentation, make them prohibitive in segmentation-free scenarios.

Another approach is based on the local features in the form of the Bag-of-Visual Words (BoVW) model [25], [26]. Rusinol *et al.* [26] present a patch-based framework where the patches are represented by a BoVW model which the local features are the SIFT descriptors. Then, they applied the Spatial Pyramid Matching method to their BoVW model to comprehend the lack of spatial information. At a next step, they normalized the BoVW descriptor with a tif-idf model and then they applied a latent semantic indexing technique. Apart of the costly, database-depended operation of creating a codebook for the BoVW model, their method further depend on the query font size, thus, making it impractical to matching word of different size due to diverse writing forms.

Llados *et al.* [25] evaluate the performance of various word descriptors, in a bag of visual words (BoVW) context, 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 reported that the

[1] http://orpheus.ee.duth.gr/cspd/



(a) (b)

Fig. 1. (a) Global diagram of the proposed word spotting framework, (b) Keypoint Detection Diagram.

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

In this paper, addressing the limitations of the aforementioned works, the proposed approach employs novel local features which are oriented for documents, namely Document-oriented Local Features (DoLF) as it is based on Connected Component Analysis. Moreover, although the proposed features use the spatial information of the current points location they are based on texture information. Finally, we propose a distance algorithm that incorporates spatial context and is employed under both segmentation-based and segmentation-free scenarios.

The main novelties of the proposed method are:

(i) Use of local features that takes in consideration the handwritten documents particularities. Therefore, it is able to detect meaningful points of the characters that reside in the documents independently of its scaling.
(ii) It provides consistency between different handwritten writing variations.
(iii) Use of the same operational pipeline in both segmentation-based and segmentation-free scenarios
(iv) Incorporation of spatial context in the local search of the matching process by integrating a near neighbor search procedure relative to the each keypoint.

The proposed paper is an considerable evolved method from the conference paper [27]. There exist fundamental differences which are given in the following:

(i) The detection of local points is based on the Connected Components (CCs) center of gravity instead of their convex hull.
(ii) The filtering of local points is more robust to scaling as it is based on CCs statistics instead on the entropy.
(iii) The local points features are calculated on a dynamic window size and not on a fixed window size.
(iv) The space normalization in matching process now incorporates the query word ratio.
(v) The proposed method works under segmentation-free scenarios as the conference paper performs only for segmentation-based.

III. Proposed Methodology

A. Introduction

The operational pipeline of the proposed word spotting framework is illustrated in Fig. 1a. It consists of two distinct steps: the Offline and the Online. At the Offline step, which is executed once, the document images are segmented to the word images for which, the proposed local features are extracted and indexed to a database. It should be noted, that the word segmentation step is disregarded under the segmentation-free scenario and local features are extracted directly from the document page image.

At the Online step, which is the only visible operation to the user, the proposed feature set is extracted from the query word image and a matching procedure is applied between the features of the query and each indexed word image. Thereafter, a ranking list of all the word images is presented to the user.

For the sake of clarity, in the case of segmentation-based scenario the ranking list comprises all the pre-segmented word images while in the case of segmentation-free scenario the ranking list comprises all detected word images, similar to the query after a complete search in the document images.

B. Document-Oriented Local Features (DoLF)

1) KeyPoints Detection: Fig. 1b shows the consecutive steps of the proposed methodology for the detection of local characteristic points (keypoints) either in a word or document image I in the case of a segmentation-based or a segmentation-free scenario, respectively.

Initially, the gradient vectors of the image I in both x-axis (I_x) and y-axis (I_y) are calculated by the convolution of the grey-level image k with the 1-D kernels $[-1, 0, 1]$ and $[-1, 0, 1]^T$, respectively. The results are shown in Fig. 2b and 2c.

Next, the I_x and I_x vectors are filtered by a high-pass filter in order to remove the noise that resides in the document background:

$$I_x^{'} = \begin{cases} 0 & if\ I_x < thr_x \\ I_x & if\ I_x \geq thr_x, \end{cases} \quad I_y^{'} = \begin{cases} 0 & if\ I_y < thr_y \\ I_y & if\ I_y \geq thr_y \end{cases} \quad (1)$$

The variables thr_x and thr_y are calculated dynamically by applying the Otsu algorithm [28] on I_x and I_x vectors, respectively, by minimizing the intra-class variance between the two clusters.

The orientation $\theta(G^I)$ (Fig. 2d) and magnitude $D(G^I)$ of the gradient vector G^I of the image I are defined as follows:

$$\theta(G^I) = tan^{-1} \cdot G^I \quad (2)$$

$$D(G^I) = \sqrt{(I_x^{'})^2 + (I_y^{'})^2}, \quad \text{where} \quad G^I = \left(I_x^{'} I_y^{'}\right) \quad (3)$$

Fig. 2d shows an example of the gradient angle image. The grey values represent zero angles. The dark colours represent negative angles while the bright colours represent positive angles.

The next step involves linear quantization of the $\theta(G^I)$. The purpose of this step is to label the changes to writing direction as these points consist of important and descriptive information. Table I shows the mapping of each $\theta(G^I)$ to corresponding quantization value and Fig. 2e shows the output for the quantization of the $\theta(G^k)$ values. Each colour represents a different quantization level.

Next, for each quantization level the Connected Components (CC) are detected, as shown at Fig. 2f-2i.

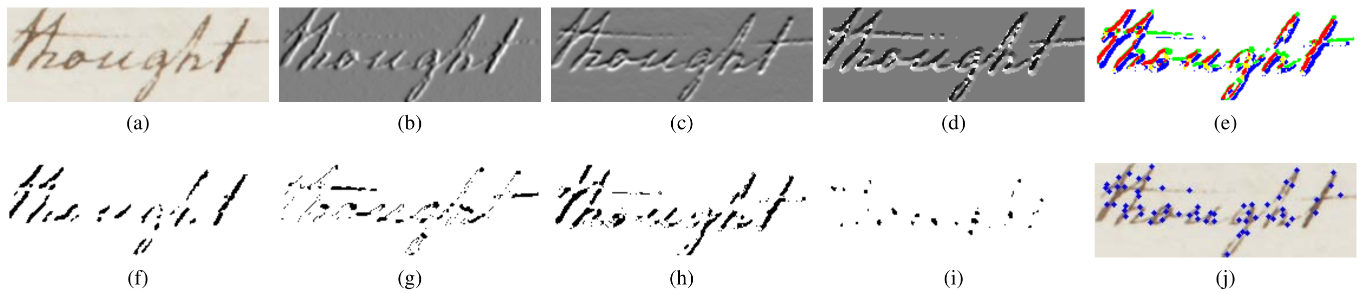


Fig. 2. The steps for the keypoints detection: (a) the original word image, (b) I_x image, (c) I_y image, (d) $\theta(G^I)$ image, (e) the quantization of the gradient angles, (f) - (i) the connected components of each quantization level, (j) the final keypoints.

TABLE I
QUANTIZATION MAPPING FOR THE GRADIENT ANGLES $\theta(G^I)$

Angle	Level
$[-180°, -90°)$	1
$[\ -90°,\ \ 0°)$	2
$[\ \ 0°,\ \ 90°)$	3
$[\ \ 90°,\ 180°]$	4

These CCs represent chunks of strokes that correspond to different writing directions between them. Then, a filtering procedure is applied to reject outliers that may affect the final result and to decrease the time expense required. Therefore, the CCs that satisfy the following equation are accepted:

$$M_{area}^{CC} - MAD_{area}^{CC} \leq CC_{area} \leq M_{area}^{CC} + MAD_{area}^{CC} \quad (4)$$

where M_{area}^{CC} is the CCs mean area size and MAD_{area}^{CC} is the CCs mean absolute deviation area size. The final local points are represented by the center of gravity of each remaining CCs. An example of these keypoints is shown in Fig. 2j.

The most meaningful parameter is the number of the quantization levels. This affects the number of the created local points as it increases the CCs.

The proposed keypoint detection method is able to detect meaningful points of the characters that reside in the documents independently of its scaling. Moreover, the linear quantization and the resulting CCs represent chunks of strokes that correspond to different writing directions between them. A subset of these CCs should be stable between different scaling and handwriting styles as some of those chucks remains the same.

In supporting that, Fig. 3 shows some qualitative results based on different handwriting variations and scales.

2) Feature Extraction: Most works using local features are based on the Scale Invariant Feature Transform (SIFT) [29] in order to describe the local information. The original application of these local features are the natural images which have many structural differences compared to document images. Firstly, the detection of the most powerful edges through pyramid scaling, creates local points between text lines. Secondly, we argue, that it is not beneficial in document images to incorporate invariant properties in the descriptor of the local points as it results in noise amplification. This is further supported by the observation stated in [17] wherein the used features which are invariant to rotation have resulted in worse performance, when compared to features that are dependent on rotation. They adhere to the observation that the features that are invariant to rotation are more sensitive to the noise.

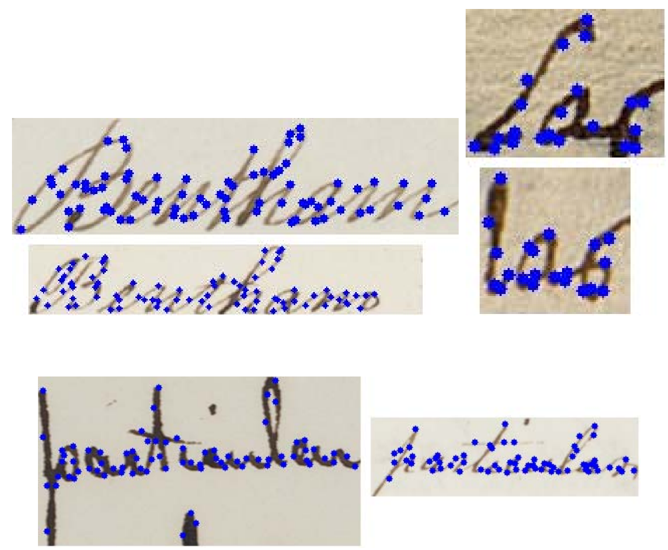


Fig. 3. Qualitative comparison of the local points detection between different writing styles and scales.

Moreover, it is worth noting that the 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 which is further verified in [25].

The features extracted at each local keypoint are based upon the orientation and the magnitude of the image I (Eq. 2 and 3). Specifically, the proposed feature vectors are calculated upon a dynamic window around the keypoint. The dynamic window aims to determine a scale-invariance window size by applying a function in a scale-space that corresponds to a series of

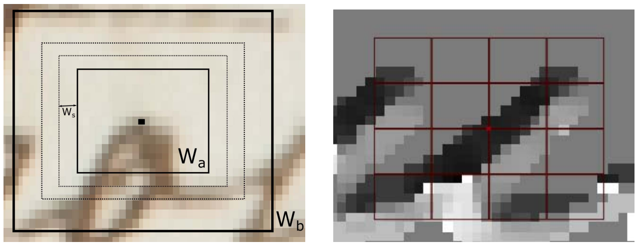


Fig. 4. (a) Representation of the dynamic window search space, (b) The window area around the keypoint.

different window sizes. Scale space theory [30], [31] allows to develop scale invariant properties by representing an image as a one-parameter function of different scaled images, the scale-space representation, and selecting the one that maximizes the above function.

Fig. 4a shows the successive window levels of the dynamic window achieved with a step equal to Ws. The W_a denote the starting window size and the W_b denote the ending window size. The optimal window is chosen based on the local maximum value of a scale-invariant function. There is a variety of functions/kernel in the bibliography such as the Laplacian or Difference of Gaussians. In the proposed approach, the mean intensity is chosen due to simplicity and speed.

When the window size is chosen, the area around the keypoint, is divided into 16 cells as shown at Fig. 4b. Each cell is represented by a 4-bin histogram (each bin corresponds to a quantization level) and each pixel accumulates a vote in the corresponding angle histogram bin. The strength of voting depends on the magnitude of the gradient vector.

Finally, all 16 histograms are concatenated in a single 64-bin histogram and normalized by its norm. Motivated by the work [29] in order to make the descriptor illumination independent, all the values above 0.2 are fixed to 0.2 and the resulting values are re-normalized again.

Although, the proposed descriptors appear to be similar to Histogram of Oriented Gradients (HOG) [32] or to Scale-invariant feature transform (SIFT) [29] descriptors, they have distinct differences:

(i) Unlike SIFT, the proposed descriptors are computed on scale-invariant key points without rotating to align orientation.

(ii) The HOG is calculated on the whole document using a dense grid without scale-invariant properties.

(iii) The proposed angle gradient quantization corresponds to the initial quantization for the local point detection method and it has only 4 levels. In contrast, the HOG histogram channels are evenly spread over 0 to 180 degrees or 0 to 360 degrees, depending on whether the gradient is unsigned or signed and SIFT uses 8 orientation levels.

IV. Word Matching

The proposed word matching method is motived from the Nearest Neighbor Search (NNS) [33] by incorporating a spatial context suitable for document images. The advantage of the proposed matching is three-fold: (i) it enables a local search instead of searching in a brute force manner, (ii) it incorporates spatial context and (iii) it is suitable under both segmentation-based or segmentation-free scenarios. In the sequel, the complete matching step will be detailed.

A. Matching in a Segmentation-Based Context

In the case of segmentation-based word spotting, the aim is to match the query features to the corresponding features of any word image in the document.

The initial stage in the matching step is a normalization which is applied for any word image including the query word image so that scale invariance is achieved.The normalized procedure comprises the following steps:

Step 1: Calculation of the mean center (c_x, c_y) of the keypoints set in a word image:

$$(c_x, c_y) = \left(\frac{\sum_{i=1}^{k} p_x^i}{k}, \frac{\sum_{i=1}^{k} p_y^i}{k} \right) \tag{5}$$

where p_x^i, p_y^i denote the location of the i^{th} keypoint and k denotes the total number of the keypoints in a word image.

Step 2: Calculation of the mean distance (D_x, D_y) of all keypoints from the mean center (c_x, c_y) is denoted as:

$$(D_x, D_y) = \left(\frac{\sum_{i=0}^{k} \left| p_x^i - c_x \right|}{k}, \frac{\sum_{i=0}^{k} \left| p_y^i - c_y \right|}{k} \right) \tag{6}$$

Step 3: Calculation of the updated normalized location for each keypoint:

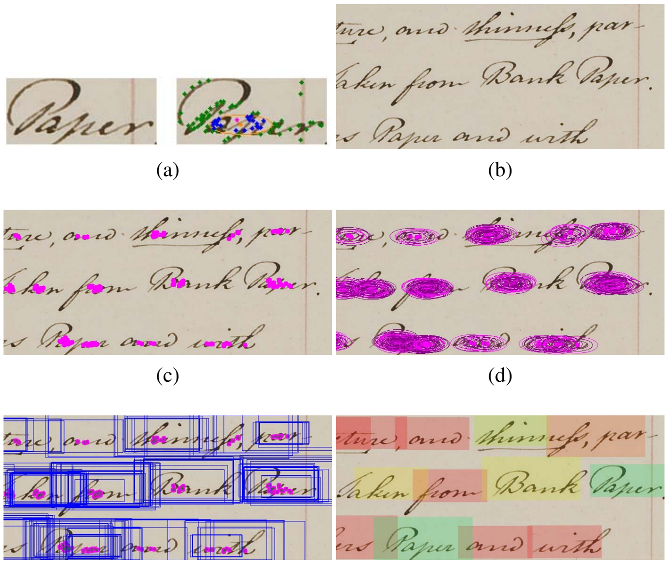
$$(p_x^{i_{new}}, p_y^{i_{new}}) = \left(\frac{p_x^i - c_x}{D_x}, \frac{p_y^i - c_y}{D_y * (w/h)} \right) \tag{7}$$

where w/h is the width to height ratio of the word image. It should be noted the the y-dimension corresponds to height.

After normalization, all word images are directly comparable due to the achieved scale invariance as shown in Fig. 6c.

In the next stage, the spatial NNS for each keypoint that resides on the query image is addressed. The spatial NNS is realized in a search area around each point. In Fig. 6c, the search area is displayed with light green. During search, if there is one or more keypoints in the proximity of the query keypoint under consideration, the Euclidean distance between their descriptors is calculated and the minimum distance is kept. This is repeated for each keypoint in the query image. The final similarity measure is the average of all the minimum distances. If there is not a local point in its proximity then the query local point is ignored.

As a final stage, the system presents to the user all the word images based on ascending sort order of the calculated similarity measure.



(a) (b) (c) (d) (e) (f)

Fig. 5. The segmentation - free operational pipeline (a) the query image, the (Qc_x, Qc_y) location (shown in magenta colour) and its nearest keypoint Qk_c (shown in orange colour), (b) the document image, (c) the candidate local points for the document coordinate origin, (d) multiple instances of $(D_x^{`}, D_y^{`})$ around each candidate coordinate origin, (e) multiple word detection (f) final result (the green colour denotes the most similar word).

B. Matching in a Segmentation-Free Context

The segmentation-free word spotting scenario using the proposed DoLFs is an extension of the procedure described for the segmentation-based word spotting scenario. Its fundamental difference is that there is no information about the potential word image on the document that should be matched to the query. In this respect, the computation of Eq. 5 - 7 does not hold. In order to circumvent this, a strategy is followed which estimates potential (c_x, c_y) and the corresponding (D_x, D_y) in the document space that comprises the following steps:

As a first step, the normalization described in the previous section using Eq.5 - 7 is applied on the query word image only, and the nearest keypoint Qk_c from (Qc_x, Qc_y) point is identified. The location of the Qk_c keypoint (Qc_x^k, Qc_y^k) is the new coordinate origin as shown in Fig. 5a.

In the second stage, a procedure is initialized in order to detect the most similar local points between the query Qk_c keypoint feature and those that reside in the document. In the achieved similarity set, we create a ranked list out of which, the top N matches are kept for further use. Fig. 5d shows the N remaining local keypoints. In our implementation, the Euclidean Distance (ED) is used, and the top N matches that kept are those that have the distance from the query keypoint Qk_c feature lower than 0.045. This threshold is experimentally defined and control the time expense of the search in the document space. Larger values does not provide any meaningful effectiveness just increase the processing cost.

In the sequel, each keypoint that belongs to the top N matches is a document candidate coordinate origin. All candidates are shown at Fig. 5c. As described in the previous section, the computation of (D_x, D_y) is required. Since, in a segmentation-free approach, there is no knowledge about the word image boundaries, we proceed in a stepwise use of multiple instances of the computed $(D_x^{`}, D_y^{`})$ at an interval that is guided by the computed query word image (QD_x, QD_y).

In particular, $D_x^{`}$ take values in the range $[\frac{2 \cdot QD_x}{3}, \frac{4 \cdot QD_x}{3}]$ and $D_y^{`}$ take values in the range $[\frac{2 \cdot QD_y}{3}, \frac{4 \cdot QD_y}{3}]$ with a step of $\frac{QD_x}{3}$ and $\frac{QD_y}{3}$, respectively. For each instance of $(D_x^{`}, D_y^{`})$, the similarity measure as described in the previous section will be used to determine the optimal $(D_x^{opt}, D_y^{opt})_{k_{p_i}}$ for each keypoint k_{p_i} that belongs to the top N matches in the document. Due to unknown word boundaries for the matching procedure, the considered local points are those that are contained in an ellipse defined by: center Qk_c and axis length (QDx, QDy) for the query case while center k_{p_i} and axis length $(D_x^{`}, D_y^{`})$ for the candidate word image case. Fig. 5a and Fig. 5d shows the aforementioned procedure.

Finally, this kind of information permits to realize a search and to proceed in a matching as described in the previous section between the keypoints of the query word image and the keypoints that are contained in the search area that is set by the optimal $(D_x^{`}, D_y^{`})$ and is centered at one of the keypoints that belongs to the top N matches.

The aforementioned procedure creates many overlapping word matches as some N keypoints reside nearby, as shown in Fig. 5e. Therefore a merging procedure is initiated. Two word rectangles are merged when the relative overlapping area is over a certain threshold. The overlapping area OA is defined as:

$$OA = \frac{A \cap B}{A \cup B} \tag{8}$$

A and B denote the two word rectangles. In our implementation the threshold is defined at 0.5. This threshold defines the overlapping sensitivity of the final results.

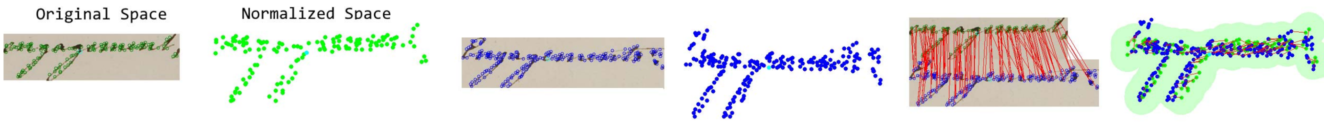
The most meaningful thresholds are the top N matches and the search area around each local point. Both, these thresholds control the time expense of the of the search area in the document space.

Finally, the system presents to the user all the word images based on ascending sort order of the calculated similarity measure as show in Fig. 5f.

V. Experimental Results

The proposed unsupervised word spotting methodology is evaluated on three datasets of handwritten documents:

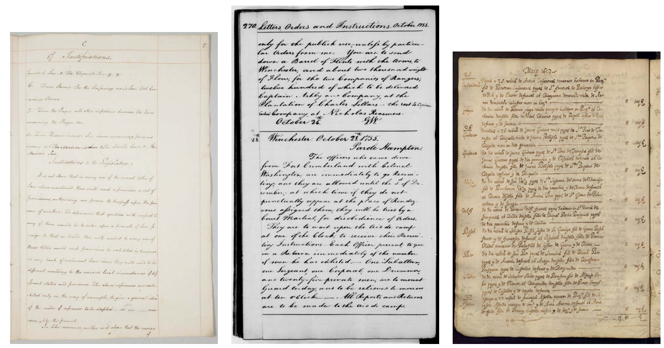
- **Bentham Dataset [34]:** It consists of high quality (approximately 3000 pixel width and 4000 pixel height) handwritten manuscripts written by Jeremy Bentham (1748-1832) himself as well as fair copies written by Bentham's secretarial staff over a period of sixty years. It contains several very difficult problems, wherein the most difficult is the word variability as a result of multiple writers. The variation of the same word is extreme and involves writing style, font size, noise as well as their combination. Fig. 7 shows some examples of these instances. It is worth to note, that we evaluate against two different document sets. The first one it denotes as Bentham-ICFHR14, it was used in ICFHR 2014 Handwritten Wordspotting Competition [8]



(a) (b) (c)

Fig. 6. Keypoint detection and correspondence at the original and the normalized space: (a) the query keypoints, (b) the word image keypoints, (c) the projection of query keypoints to the word image: the red lines connect the matched local points, the green area (right image) is the local proximity area of the nearest neighbour search in the normalized space.

Fig. 7. Type of word variations met in the Bentham Dataset for the words 'England' and 'Embezzlement'.



(a) (b) (c)

Fig. 8. Representative document images from (a) Bentham Dataset, (b) Washington Dataset, (c) Barcelona Historical Handwritten Marriages Dataset.

and contains 50 documents. The second one it denotes as Bentham-ICDAR15, it was used in ICDAR2015 Competition on Keyword Spotting for Handwritten Documents [9] and contains 70 documents. The datasets and their corresponding query sets are freely available in their associated websites.[2]

- **Washington Dataset [12]:** It consists of document images from George Washington Collection of the Library of Congress (namely 3000300 - 3000309) [12]. The dataset and the corresponding query set is available to their website.[3] It is worth to note, that we evaluate against two different experimental setups. One, that it uses 10 good quality pages and it has 2381 queries and another one that consists of 20 handwritten pages with a total of 4860 words.
- **Barcelona Historical Handwritten Marriages Dataset (BH2M) [35]:** It consists of 40 images of historical handwritten marriages records stored in the archives of Barcelona cathedral. This book was written between 1617 and 1619 by a single writer in old Catalan. It is available at their website.[4]

Fig. 8 shows some representative document images from each of those datasets.

The measures employed in the performance evaluation of the proposed segmentation-based algorithm are the Precision at the 5 Top Retrieved words (P@5) and the Mean Average Precision (MAP). To further detail the metrics, let define Precision and P@k as follows:

$$P@k = \frac{|\{relevant\ words\} \cap \{k\ retrieved\ words\}|}{|\{k\ retrieved\ words\}|} \quad (9)$$

Precision is the fraction of retrieved words that are relevant to the search, while in the case that precision should be determined for the k top retrieved words, P@k is computed. In particular, in the proposed evaluation, P@5 is used which 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.

The second metric used in the proposed evaluation is the Mean Average Precision (MAP) which is a typical measure for the performance of information retrieval systems [36]. It is implemented from the Text Retrieval Conference (TREC) community by the National Institute of Standards and Technology (NIST). The above metric is defined as the average of the precision value obtained after each relevant word is retrieved:

$$AP = \frac{\sum_{k=1}^{n} (P@k \times rel(k))}{\{relevant\ words\}} \quad (10)$$

where:

$$rel(k) = \begin{cases} 1, & if\ word\ at\ rank\ k\ is\ relevant \\ 0, & if\ word\ at\ rank\ k\ is\ not\ relevant \end{cases} \quad (11)$$

Both selected evaluation metrics are well known and established for retrieval. The P@5 is used towards evaluating precision-oriented performances and MAP for general retrieval performance.

A. Segmentation-Based Evaluation Performance

In the segmentation-based context is assumed that there is an outcome of a word image segmentation method. As the scope of the proposed word spotting method is in the local features and its accompanied matching process, there will be no discussion about any specific methodology used for the segmentation process. In particular, for the experiments,

[2]http://vc.ee.duth.gr/h-kws2014 and http://transcriptorium.eu/~icdar15kws/
[3]http://ciir.cs.umass.edu/downloads/old/data_sets.html
[4]http://dag.cvc.uab.es/the-historical-marriages-database

TABLE II

EXPERIMENTAL RESULTS FOR SEGMENTATION-BASED IN BENTHAM DATASET

ICFHR14			ICDAR15		
Method	*P@5*	*MAP*	*Method*	*P@5*	*MAP*
Kovalchuk et al. [19]	0.738	0.524	PRG Group	0.460	0.424
Almazan et al. [20]	0.724	0.513	CVC Group [38]	0.343	0.300
Howe et al. [39]	0.718	0.462	Zagoris et al. [27]	0.268	0.217
Zagoris et al. [27]	0.623	0.393	CSPD [13]	0.224	0.193
CSPD [13]	0.525	0.341	**Proposed Method**	**0.501**	**0.440**
Proposed Method	**0.788**	**0.600**			

TABLE III

EXPERIMENTAL RESULTS FOR SEGMENTATION-BASED IN WASHINGTON DATASET

Setup A		Setup B	
Method	*MAP*	*Method*	*MAP*
Zagoris et al. [27]	0.620	Kovalchuk et al. [19]	0.663
CSPD [13]	0.613	Zagoris et al. [27]	0.405
Rath and Manmatha [11]	0.409	CSPD [13]	0.401
Rothfeder et al. [40]	0.362	Wang et al. [41]	0.175
Proposed Method	**0.743**	**Proposed Method**	**0.692**

TABLE IV

EXPERIMENTAL RESULTS FOR BH2M DATASET

Segmentation-based		***Segmentation-free***	
Method	*MAP*	*Method*	*MAP*
Zagoris et al. [27]	0.407	HOG + EWS [43]	0.513
CSPD [13]	0.358	**Proposed Method**	**0.530**
Vinciarelli + DTW [44]	0.315		
Proposed Method	**0.607**		

TABLE V

EXPERIMENTAL RESULTS USING THE OUTPUT OF AN SEGMENTATION METHOD IN BENTHAM-ICFHR14 DATASET

Method	P@5	MAP
Segmentation-based KWS (ground truth)	0.788	0.600
Segmentation-based KWS (imperfect segmentation)	0.694	0.519
Segmentation-free	0.710	0.517

the word image segmentation information is taken from the ground truth corpora.

Initially, the proposed algorithm is evaluated against the ICFHR 2014 Handwritten Competition [8]. The query set size is 320 and it contains word images with length greater than 6 and frequency greater than 5. Table II shows the results for the segmentation-based track (Track I). Our proposed method provides considerable better performance than the other contesters. It worth to note that Almazan *et al.* method [20] while it is based on SIFT local features, it uses a machine learning procedure (Gaussian mixture model) and therefore it requires training data.

Moreover, it is evaluated taking into account the ICDAR2015 Competition on Keyword Spotting for Handwritten Documents [9]. The query set consists of 243 distinct words of different lengths (6 to 15 characters). Each of these words is represented by 6 or less different word images, making a total of 1421 query images. Table II presents the results for segmentation-based Track I.A. The PRG (Pattern Recognition Group, TU Dortmund University, Germany) group created a codebook having extracted the SIFT descriptors from all test words and the CVC (Computer Vision Center, Universitat Autnoma de Barcelona, Spain) group uses Integral Histogram of Gradients (IHOG) in a Bag-of-Visual-Words framework. On contrary, although our proposed approach does not use any dataset-specific optimization stage, it provides the best performance.

Next, we evaluated the word spotting performance in the Washington dataset. Although it is widely used, there is not a standard experimental setup, and each work adapts it to the needs of their proposed algorithm. In order to circumvent this problem, we choose to evaluate against the experimental setup that appears in [41]. It is based on two different configuration setups:

- **Setup A:** It uses 10 good quality pages (pages 3000300-3090309) and it has 2381 queries.
- **Setup B:** This dataset consists of 20 handwritten pages with a total of 4860 words. It use as queries only words which have at least ten occurrences and with 3 or more characters.

Table III shows the evaluation results against similar word spotting methods (without any training data or any learning-based algorithm). Our proposed method achieves the best performance.

Next, the proposed method is evaluated in BH2M dataset. The query set comprises of 5170 word image queries. Table IV shows the evaluation results where the proposed approach ranks first.

Finally, to investigate the performance of the segmentation-based word spotting under a scenario that involves an imperfect word segmentation, an artificial random error is introduced to ground truth word information. In particular, the word width or height as well as the have be randomly reduced or increased of the total width or height by 20%, respectively. The segmentation-based KWS evaluation results using the above imperfect word segmentation output are shown in Table V. The evaluation dataset is the Bentahm-ICFHR14 from the segmentation-free track. Moreover, the segmentation-free KWS evaluation results are added for reference.

B. Segmentation-Free Evaluation Performance

First of all, it is worth to note that in the experiments for the segmentation-free case, a result bounding box may not match exactly with the word bounding box from ground-truth corpora. Thus, a correct match is registered when the relative overlapping area is over a certain threshold. For the sake of consistency, in every segmentation-free experiment the overlapping area is defined as: $OA = A \cap B / A \cup B$

where OA is the overlapping area, A is the result bounding box and B is the ground-truth box. For all the experiments the threshold is defined at 0.5.

TABLE VI

EXPERIMENTAL RESULTS FOR EGMENTATION-FREE FOR BENTHAM DATASET

ICFHR14			ICDAR15		
Method	*P@5*	*MAP*	*Method*	*P@5*	*MAP*
Kovalchuk et al. [19]	0.617	0.423	PRG Group	0.376	0.293
Howe et al. [39]	0.608	0.409	CVC Group [38]	0.150	0.116
Pantke [45]	0.610	0.397	**Proposed Method**	**0.387**	**0.326**
Leydier [17]	0.353	0.221			
Proposed Method	**0.710**	**0.517**			

TABLE VII

EXPERIMENTAL RESULTS FOR LOCAL FEATURES IN BENTHAM-ICFHR14 DATASET

Method	MAP (Segmentation - based)	MAP (Segmentation- free)
SIFT [29]	0.331	0.115
SURF [46]	0.279	0.106
BRISK [47]	0.189	0.035
ORB [48]	0.334	0.098
KAZE [49]	0.399	0.283
Proposed Method	**0.600**	**0.517**

TABLE VIII

EXECUTION TIME FOR BENTHAM ICFHR14 BENTHAM DATASET

Method	Aver. Retriev. Time per Query	Extraction Time per Doc.
Segmentation-based KWS	0.61 sec	6.07 sec
Segmentation-free KWS	7.19 sec	10.76 sec

Table VI shows the results for the segmentation-free (Track II) from ICFHR 2014 Handwritten Competition which involves 290 queries and the results for the ICDAR 2015 Competition segmentation-free TRACK I.B. that involves 1421 queries. It is worth to note that the PRG Group in order to bypass the lack of information about word location, applies a word segmentation algorithm and then it uses the same procedure from the segmentation-based scenario.

Finally, Table IV shows the results for the BH2M dataset under the segmentation-free scenario. It is worth to note that the EWS [42] method uses the exemplar-SVM framework [44], and relies on a sliding-window search to retrieve the document regions that are likely to contain the query word.The proposed method, although it does not uses any segmentation algorithm or any machine learning procedure displays an increase in effectiveness in all the segmentation-free experiments.

C. Local Points Evaluation Performance

In order to showcase the performance gains of our proposed local features, they were replaced with other local features such as: (i) SIFT [29], (ii) SURF [45], (iii) BRISK [46], (iv) ORB [47], (v) KAZE [48]. The proposed matching algorithm is kept the same for all features. Table shows VII the experimental results under both segmentation-based and segmentation-free scenarios for the Bentham-ICFHR14 Dataset. It shows that the proposed local features provide considerable increase in word spotting effectiveness under both scenarios.

D. Execution Time

The execution time of the proposed method is presented in Table VIII for the ICFHR14 Bentham Dataset. The results show one major disadvantage of segmentation-free which is the considerable increase of the retrieval time.

VI. CONCLUSION

In this work, novel local features are proposed driven by the challenges presented in historical handwritten word spotting scenarios. It shows considerable effectiveness against other local features under two different word spotting scenarios: segmentation - based and segmentation - free. Moreover, a matching procedure was presented based on Nearest Neighbour, that augments performance in terms of effectiveness and efficiency incorporating spatial context. It is proven that the proposed framework achieves better performance after a consistent evaluation against 4 datasets and 13 different stage of the art methods under two different word spotting scenarios (segmentation-based and segmentation-free). Finally, an implementation of the proposed word spotting method as a recommender system to a transcription process is available at http://vc.ee.duth.gr/ws [50].

REFERENCES

[1] R. Manmatha, C. Han, and E. M. Riseman, "Word spotting: A new approach to indexing handwriting," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Proc. (CVPR)*, Jun. 1996, pp. 631–637.

[2] E. Hassan, S. Chaudhury, and M. Gopal, "Word shape descriptor-based document image indexing: A new dbh-based approach," *Int. J. Document Anal. Recognit.*, vol. 16, no. 3, pp. 227–246, 2013.

[3] K. Zagoris, K. Ergina, and N. Papamarkos, "A document image retrieval system," *Eng. Appl. Artif. Intell.*, vol. 23, no. 6, pp. 872–879, 2010.

[4] A. L. Kesidis, E. Galiotou, B. Gatos, and I. Pratikakis, "A word spotting framework for historical machine-printed documents," *Int. J. Document Anal. Recognit.*, vol. 14, no. 2, pp. 131–144, 2011.

[5] T. Konidaris, B. Gatos, S. Perantonis, and A. Kesidis, "Keyword matching in historical machine-printed documents using synthetic data, word portions and dynamic time warping," in *Proc. 8th Int. Workshop Document Anal. Syst. (DAS)*, 2008, pp. 539–545.

[6] F. Zirari, A. Ennaji, S. Nicolas, and D. Mammass, "A methodology to spot words in historical arabic documents," in *Proc. ACS Int. Conf. Comput. Syst. Appl. (AICCSA)*, 2013, pp. 1–4.

[7] K. Khurshid, C. Faure, and N. Vincent, "Word spotting in historical printed documents using shape and sequence comparisons," *Pattern Recognit.*, vol. 45, no. 7, pp. 2598–2609, 2012.

[8] I. Pratikakis, K. Zagoris, B. Gatos, G. Louloudis, and N. Stamatopoulos, "ICFHR 2014 competition on handwritten keyword spotting (H-KWS 2014)," in *Proc. 14th Int. Conf. Frontiers Handwriting Recognit. (ICFHR)*, Sep. 2014, pp. 814–819.

[9] J. Puigcerver, A. H. Toselli, and E. Vidal, "ICDAR2015 competition on keyword spotting for handwritten documents," in *Proc. ICDAR*, Aug. 2015, pp. 63–67.

[10] T. M. Rath and R. Manmatha, "Features for word spotting in historical manuscripts," in *Proc. 7th Int. Conf. Document Anal. Recognit.*, 2003, pp. 218–222.

[11] T. M. Rath and R. Manmatha, "Word image matching using dynamic time warping," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 2003, p. II–521.

[12] V. Lavrenko, T. M. Rath, and R. Manmatha, "Holistic word recognition for handwritten historical documents," in *Proc. 1st Int. Workshop Document Image Anal. Libraries*, 2004, pp. 278–287.

[13] K. Zagoris, K. Ergina, and N. Papamarkos, "Image retrieval systems based on compact shape descriptor and relevance feedback information," *J. Vis. Commun. Image Represent.*, vol. 22, no. 5, pp. 378–390, 2011.

[14] D. E. Gustafson and W. C. Kessel, "Fuzzy clustering with a fuzzy covariance matrix," in *Proc. IEEE Conf. Decision Control Including 17th Symp. Adapt. Process.*, vol. 17. Jan. 1978, pp. 761–766.

[15] S. Srihari, H. Srinivasan, P. Babu, and C. Bhole, “Handwritten arabic word spotting using the cedarabic document analysis system,” in *Proc. Symp. Document Image Understand. Technol. (SDIUT)*, 2005, pp. 123–132.
[16] Y. Leydier, F. Lebourgeois, and H. Emptoz, “Text search for medieval manuscript images,” *Pattern Recognit.*, vol. 40, no. 12, pp. 3552–3567, 2007.
[17] Y. Leydier, A. Ouji, F. LeBourgeois, and H. Emptoz, “Towards an Omnilingual word retrieval system for ancient manuscripts,” *Pattern Recognit.*, vol. 42, no. 9, pp. 2089–2105, 2009.
[18] B. Gatos and I. Pratikakis, “Segmentation-free word spotting in historical printed documents,” in *Proc. 10th Int. Conf. Document Anal. Recognit. (ICDAR)*, Jul. 2009, pp. 271–275.
[19] A. Kovalchuk, L. Wolf, and N. Dershowitz, “A simple and fast word spotting method,” in *Proc. 14th Int. Conf. Frontiers Handwriting Recognit. (ICFHR)*, Sep. 2014, pp. 3–8.
[20] J. Almazan, A. Gordo, A. Fornes, and E. Valveny, “Word spotting and recognition with embedded attributes,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 12, pp. 2552–2566, Dec. 2014.
[21] A. Fischer, A. Keller, V. Frinken, and H. Bunke, “Lexicon-free handwritten word spotting using character HMMs,” *Pattern Recognit. Lett.*, vol. 33, no. 7, pp. 934–942, 2012.
[22] B. Huang, C. Du, Y. Zhang, and M.-T. Kechadi, “A hybrid HMM-SVM method for online handwriting symbol recognition,” in *Proc. 6th Int. Conf. Intell. Syst. Design Appl. (ISDA)*, vol. 1. Oct. 2006, pp. 887–891.
[23] S. España-Boquera, M. Castro-Bleda, J. Gorbe-Moya, and F. Zamora-Martinez, “Improving offline handwritten text recognition with hybrid HMM/ANN models,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 4, pp. 767–779, Apr. 2011.
[24] S. Thomas, C. Chatelain, L. Heutte, T. Paquet, and Y. Kessentini, “A Deep HMM model for multiple keywords spotting in handwritten documents,” *Pattern Anal. Appl.*, vol. 18, no. 4, pp. 1003–1015, 2015. [Online]. Available: http://dx.doi.org/10.1007/s10044-014-0433-3
[25] J. Lladós, M. Rusinol, A. Fornes, D. Fernandez, and A. Dutta, “On the influence of word representations for handwritten word spotting in historical documents,” *Int. J. Pattern Recognit. Artif. Intell.*, vol. 26, no. 5, p. 1263002, 2012.
[26] M. Rusinol, D. Aldavert, R. Toledo, and J. Lladós, “Browsing heterogeneous document collections by a segmentation-free word spotting method,” in *Proc. ICDAR*, Sep. 2011, pp. 63–67.
[27] K. Zagoris, I. Pratikakis, and B. Gatos, “Segmentation-based historical handwritten word spotting using document-specific local features,” in *Proc. 14th Int. Conf. Frontiers Handwriting Recognit. (ICFHR)*, Sep. 2014, pp. 9–14.
[28] N. Otsu, “A threshold selection method from gray-level histograms,” *Automatica*, vol. 11, nos. 285–296, pp. 23–27, 1975.
[29] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[30] T. Lindeberg, “Feature detection with automatic scale selection,” *Int. J. Comput. Vis.*, vol. 30, no. 2, pp. 79–116, 1998.
[31] T. Lindeberg, *Scale-Space Theory in Computer Vision*, vol. 256. New York, NY, USA: Springer, 2013.
[32] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, vol. 1. no. 1, pp. 886–893.
[33] N. S. Altman, “An introduction to kernel and nearest-neighbor nonparametric regression,” *Amer. Statist.*, vol. 46, no. 3, pp. 175–185, 1992.
[34] D. G. Long and A. T. Milne, *The Manuscripts of Jeremy Bentham : A Chronological Index to the Collection in the Library of University College London*, 1st ed. London, U.K.: Univ. London, 1981.
[35] D. Fernández-Mota, J. Almazán, N. Cirera, A. Fornés, and J. Lladós, “BH2M: The Barcelona historical, handwritten marriages database,” in *Proc. 22nd Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2014, pp. 256–261.
[36] S. A. Chatzichristofis, K. Zagoris, and A. Arampatzis, “The TREC files: The (ground) truth is out there,” in *Proc. 34th Int. ACM SIGIR Conf. Res. Develop. Inf. (SIGIR)*, New York, NY, USA, 2011, pp. 1289–1290.
[37] D. Aldavert, M. Rusinol, R. Toledo, and J. Lladós, “Integrating visual and textual cues for query-by-string word spotting,” in *Proc. 12th Int. Conf. Document Anal. Recognit. (ICDAR)*, Aug. 2013, pp. 511–515.
[38] N. R. Howe, “Part-structured Inkball models for one-shot handwritten word spotting,” in *Proc. 12th Int. Conf. Document Anal. Recognit. (ICDAR)*, Aug. 2013, pp. 582–586.
[39] J. L. Rothfeder, S. Feng, and T. M. Rath, “Using corner feature correspondences to rank word images by similarity,” in *Proc. Conf. Comput. Vis. Pattern Recognit. Workshop (CVPRW)*, vol. 3. Jun. 2003, p. 30.
[40] P. Wang, V. Eglin, C. Garcia, C. Largeron, J. Lladós, and A. Fornes, “A novel learning-free word spotting approach based on graph representation,” in *Proc. 11th IAPR Int. Workshop Document Anal. Syst. (DAS)*, Apr. 2014, pp. 207–211.
[41] D. Aldavert, M. Rusinol, R. Toledo, and J. Lladós, “A study of bag-of-visual-words representations for handwritten keyword spotting,” in *Proc. Int. J. Document Anal. Recognit. (IJDAR)*, vol. 18, no. 3, pp. 1–12, 2015. [Online]. Available: http://dx.doi.org/10.1007/s10032-015-0245-z
[42] J. Almazán, A. Gordo, A. Fornés, and E. Valveny, “Efficient exemplar word spotting,” *BMVC*, vol. 1, no. 2, p. 3, 2012.
[43] A. Vinciarelli and S. Bengio, “Off-line cursive word recognition using continuous density HMMs trained with pca or ica features,” in *Proc. Int. Conf. Pattern Recognit. (ICPR)*, 2002, pp. 493–498.
[44] T. Malisiewicz, A. Gupta, and A. A. Efros, “Ensemble of exemplar-SVMs for object detection and beyond,” in *Proc. ICCV*, Nov. 2011, pp. 89–96.
[45] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-up robust features (SURF),” *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.
[46] S. Leutenegger, M. Chli, and R. Siegwart, “BRISK: Binary robust invariant scalable keypoints,” in *Proc. IEEE Int. Conf. Comput.*, Nov. 2011, pp. 2548–2555.
[47] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “ORB: An efficient alternative to SIFT or SURF,” in *Proc. IEEE Int. Conf. Comput.*, Nov. 2011, pp. 2564–2571.
[48] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, “Kaze features,” in *Proc. 12th Eur. Conf. Comput. Vis.-ECCV*, vol. 6. Berlin, Heidelberg, 2012, pp. 214–227.
[49] W. Pantke *et al.*, “HADARA—A software system for semi-automatic processing of historical handwritten Arabic documents,” in *Proc. Arch. Conf.*, Washington, DC, USA, Apr. 2013, pp. 161–166.
[50] K. Zagoris, I. Pratikakis, and B. Gatos, “A framework for efficient transcription of historical documents using keyword spotting,” in *Proc. 3rd Int. Workshop Historical Document Imag. Process. (HIP)*, Aug. 2015, pp. 9–14.

Konstantinos Zagoris, photograph and biography not available at the time of publication.

Ioannis Pratikakis, photograph and biography not available at the time of publication.

Basilis Gatos, photograph and biography not available at the time of publication.