2017-07-21

Handwritting: keyword spotting The Query by Example (QbE) case

Barlas, Georgios

Nova Science Publishers, Inc

Chapter 11

Handwritten keyword spotting The Query by Example (QbE) case

Georgios Barlas, Konstantinos Zagoris and Ioannis Pratikakis

Georgios Barlas, Konstantinos Zagoris and Ioannis Pratikakis
Department of Electrical and Computer Engineering
Democritus University of Thrace
67100 Xanthi, Greece

Corresponding Author: ipratika@ee.duth.gr

Introduction

The traditional approach in document indexing usually involves an Optical Character Recognition (OCR) step. Although OCR performs well in modern printed documents and documents of high quality printing, in the case of handwritten documents OCR, several factors affect the final performance like intense degradation, paper-positioning variations (skew, translations, etc.) and writing styles variety.

Handwritten word spotting has attracted the attention of the research community in the field of document image analysis and recognition since it appears to be a feasible solution for indexing and retrieval of handwritten documents in the case that OCR-based methods fail to deliver satisfactory results.

Handwritten keyword spotting (KWS) is the task of retrieving all instances of a given query word in handwritten document image collections without involving a traditional OCR step. There exist two basic variations for KWS approaches: (a) the Query by Example case (QbE) where the query is a word image and (b) the Query by String case (QbS) where, as the name implies, the query is a string. The study presented in this chapter will focus on the QbE approach.

For a better understanding, QbE methods will be presented taking into account two different perspectives which relate to the use of segmentation and learning. The segmentation-

based methods are divided into 2 subcategories based upon the segmented entity which could be either the word image or the textline. They are strongly dependent on the segmentation step, so that to compare different methods regardless of segmentation errors, many researchers do not implement a segmentation method but they use datasets where the segments are given.

In the case of segmentation-free methods the whole image is tested against similarities between the query image and the patches of the document image without segmenting it at any level. The methods of this class, on the one hand bypass the step of segmentation but on the other hand they cannot avoid searching for the words in parts of the image that may not contain text. Therefore, segmentation-free methods avoid failures due to bad segmentation but the running time increases considerably. It is worth-mentioning that the methods of this class are not the trend.

Training-based methods are those that require training data at a particular stage of the process. A common problem in these methods is the availability of training data. Furthermore, an extra weakness is that to apply such a method to a new word, usually ground truthing work is required to obtain training data, which is quite time consuming and often it has to be done totally manual.

Training - free are methods that as the name implies do not include any training stage in the operational KWS pipeline. The training - free methods can be applied directly to new word although, they usually require a particular configuration to be effective in the corresponding text.

This chapter is structured as follows: Section "Segmentation-based Context" will present the KWS methodologies that operate in a segmentation-based context wherein methods based on training and methods that are independent of any training involvement will be detailed. Both variations will be separately reviewed depending on the type of segmentation which is used. In Section "Segmentation - Free Context", methodologies that do account for a segmentation will be discussed with a particular focus on the use or not of training. Section "Experimental Datasets and Evaluation Metrics" deals with an overview of the current efforts for performance evaluation and a brief description of datasets that were used in QbE KWS, while the Section "Conclusive Remarks" is dedicated to a fruitful discussion which aims to identify the current trends of the QbE KWS.

Segmentation-based Context

In this section, segmentation-based methods are presented. Segmentation-based methods have been categorized into training-based and training-free approach. Each category is then subdivided into word image segmentation and textline segmentation context.

Methods Based on Training

Word Image Segmentation Context

In the work of Rodríguez-Serrano and Perronnin (2009), the method is based on a Semi-Continuous - Hidden Markov Model (SC-HMM) coupled with a Gaussian Mixture Model (GMM). SC-HMM is able to learn from a small set of samples. A segmentation algorithm

extracts sub-images that potentially represent words, employing state-of-the-art techniques based on projection profiles and clustering of gap distances. Then, a simple classifier using holistic features (per column, pixel count, Local Gradient Histogram (LGH)) is employed for performing a first rejection pass. The non-rejected word images are normalized with respect to slant, skew and text height, using standard techniques. Then, for each normalized word image, LGH features are computed by moving a window from left to right over the image and feature vectors are extracted at each position to build the feature vector sequence. Finally, using SC-HMM with GMM, a score is assigned to each feature vector sequence which is used to attribute the similarity with the query using a threshold. An overview of the methodology is shown in Figure 11.1.

The same framework was used in Rodríguez-Serrano et al. (2010) modified so that writers adaptation is achieved. For this purpose, a statistical adaptation technique was applied to change some of GMMs parameters at each document. Furthermore, SC-HMM was used in Rodríguez-Serrano and Perronnin (2012) to enrich features extraction, since in a left-to-right HMM, the states are ordered and the weights of the SC-HMMs can be viewed as a sequence of vectors. The distance between these vectors is computed using the Dynamic Time Warping (DTW) wherein Bhattacharyya measure is being used as local similarity. The use of SC-HMM in an unsupervised context was presented in Rodriguez-Serrano and Perronnin (2012) where character examples of existing fonts were used to create the training set.

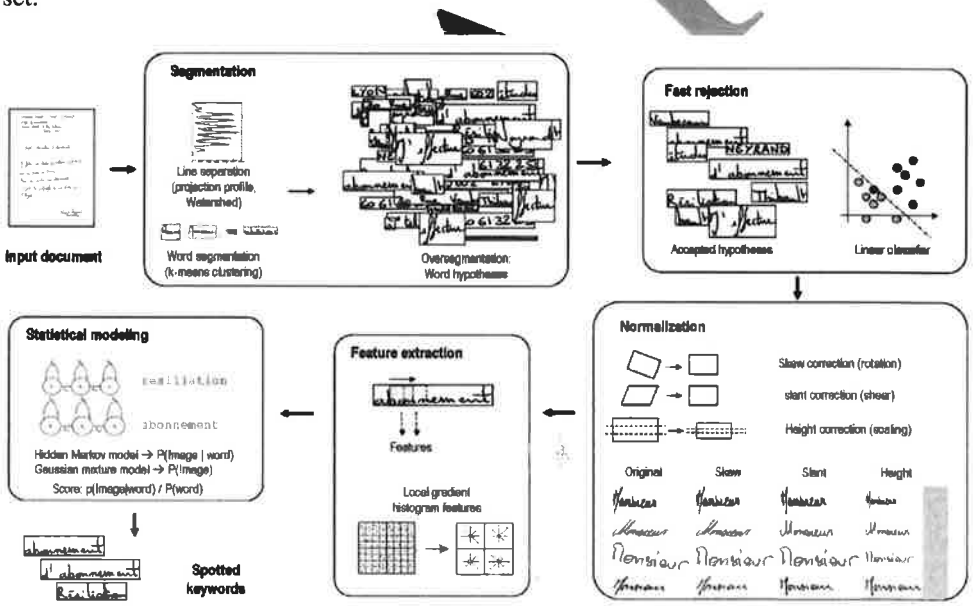


Figure 11.1. Overview of the word-spotting system presented in (Rodríguez-Serrano and Perronnin, 2009).

In the work in Almazán et al. (2012a), a preprocessing stage is initially applied using margins removal and anisotropic Gaussian filtering. Then, binarization and word segmentation is applied. In the core methodology, they use a hashing strategy based on Loci features

to prune word images and limit the candidate locations. A discriminative model is then applied to those locations. The discriminative learning relies upon a Support Vector Machine (SVM) which sets the weights on the appearance features Histogram of Oriented Gradients (HOG) to compute the final similarity.

Almazan et al. (2013) created a method that is both QbE and QbS and addresses multi-writer WS. The QbE pipeline is based on Fisher Vectors (FV) computed over Scale-Invariant Feature Transform SIFT descriptors. Then the FV are used to feed an SVM to get the attribute scores. They report that any encoding method and classification algorithm that transforms the input image into attribute scores could be used to replace them, but they chose SVMs and FVs for simplicity and effectiveness.

The study in Fernández et al. (2013) uses a previous method (Fernández et al., 2011) which is extended in a way that the syntactic context in the document is used to infer context. To achieve this, Markov Logic Networks (MLN) are used that are trained with specific rules. The MLN can be considered as a collection of first order rules to each of which it is assigned a real number, the weight. Each rule represents the rule in the domain, while the weights indicate the strength of the rule.

The work in Aldavert et al. (2015) uses the Bag-of-Visual-Words (BoVW) model to WS. The authors divide the procedure of creating BoVW in four basic steps: sampling, description, encoding and pooling. In particular, they sampled densely the word images using a fixed step at different scales. The description is derived from the HOG descriptor. To encode the descriptors and create the codebook, the Locality-constrained Linear Coding (LLC) algorithm was used. Finally, at pooling step, the Spatial Pyramid Matching (SPM) technique applied so that spatial information could be used.

In the work at Sharma et al. (2015), they made experiments with Convolutional Neural Networks (CNN). Only the classification layers were retrained to address the problem. The CNN extracted a 4096-d feature for each word image which was achieved after discarding the last fully connected layer and considering the activation of the units in the second fully connected layer. For matching, standard Lp norms have been used.

Textline Segmentation Context

The system presented in Keaton et al. (1997) is composed by several modules. The "focus-of attention" module concerns a cross-correlation testing between the query image and the document for finding the candidate locations. The "preprocessing" module that consists of estimation of word image zones at the upper, middle and lower level, filtering of stray marks and skeletonization. The "feature extraction" module that concerns profile encoding (20 Discrete Cosine Transform (DCT) coefficients from the profile extracted at each of the three zones and cavity encoding which takes into account 2D spatial arrangement, as well as the descender, and ascender information leading to a graph. Both encoding features are combined to a new graph that contains the type, size and relative location of each feature, which is considered as keyword signature graph. Finally, the keyword signature matching is addressed by two distinct comparisons. First, a comparison is employed between the profile encoding DCT coefficients wherein the resulting comparison is incorporated into the graph as additional feature. In the sequel, the keyword signature graphs are compared with probabilistic graph matching based on Bayesian evidential reasoning.

Training - Free Methods

Word Image Segmentation Context

In the paper Manmatha et al. (1996) the term "word spotting" was introduced for handwritten documents as analogous to "word spotting" in speech processing. It was applied on scanned greylevel document images. The steps of the algorithm are gaussian filtering, subsampling to reduce image by half, binarization by thresholding and segmentation into words. Then follows pruning taking into account (i) the aspect ratio and (ii) the size using predefined thresholds. Finally, at matching stage, two different matching algorithms were used based on the standard and the affine-corrected (SLH algorithm) Euclidean distance, respectively.

In the sequel, Rath and Manmatha (2003b), was motivated by Kolcz et al. (2000) that have used Dynamic Time Warping for matching in combination with a textline segmentation method. In this approach, the textine segmentation has been replaced by a word image segmentation. In the work of Rath and Manmatha (2003a), the feature set used for the experiments was extended to 11 distinct features (4 projection profiles, 2 word profiles, background ink transitions, graylevel variance and 3 Gaussian smoothing variation features) which were used as single or combined ones for matching with DTW. Nevertheless, the more recent work in Rath and Manmatha (2007) suggested only three distinct features, namely, projection profiles, word profiles and background ink transitions which were optimally matched when using DTW.

A study based on contours of words was presented in Adamek et al. (2007). They start with local binarization of the image. To achieve smoothness at word outlines, they preprocess the image applying morphological filtering. After binarization a word may split to more than one connected component. To estimate the exact position of the word in the word image a process based on horizontal and vertical projection histogram and a fixed thresholding is applied. Then, after applying a series of heuristics rules, the connected components of the word image are linked together to create a single component from which the contour is extracted. They used Multiscale Convexity Concavity (MCC) representation for the contour. MCC calculates the convexity and concavity along contour at different scales to create 2D matrix were rows correspond to scale level and columns to convexity or concavity. To measure the matching between the contours the DTW algorithm were used. The distance matrix of DTW is constructed by storing the distance between a pair of contour points corresponding to the row and column of the MCC representation. The final dissimilarity between contours is the normalized optimal path of the matrix. An alternative method that was tested is the MCC-DCT where the 1D DCT is applied at MCC representation matrix and the coefficients of DCT are combined to the final dissimilarity.

The work in Bhardwaj et al. (2008) presented an algorithm based on moment functions. In the initial stage, they used horizontal and vertical profiles to segment the document into lines and words, respectively. High order (up to 7) moment functions were used to extract features and indexing each word image. They used cosine similarity metric for matching and relevance feedback to improve the results.

A shape descriptor, the Compact Shape Portrayal Descriptor (CSPD) was presented in Zagoris et al. (2011) which requires only 123 bits per word image. CSPD is based on five (5) distinct features: (i) width height ratio, and the DCT coefficients of (ii) vertical projection,

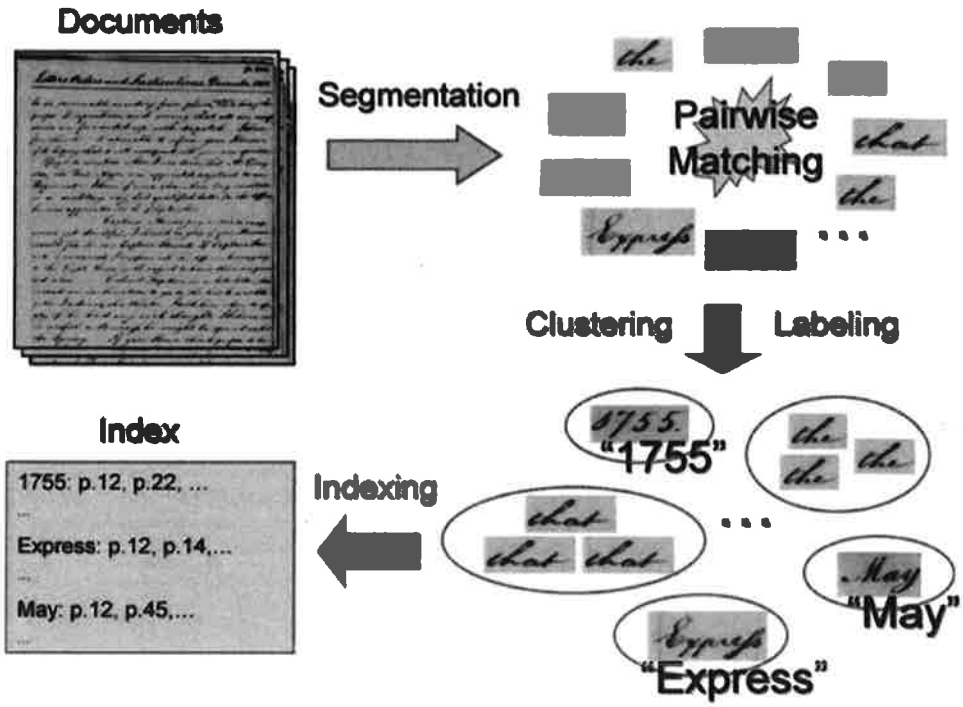


Figure 11.2. An illustration of the word spotting pipeline presented in (Rath and Manmatha, 2007).

(iii) horizontal projection, (iv) top shape projection and (v) bottom shape projection. For each feature Gustafson-Kessel fuzzy algorithm is applied for quantization to reduce the stored size of each descriptor. To refine the results, relevance feedback is applied. The user selects the best result and a training set is created. This training set is used to train an SVM to correct the results. The similarity measure that was used is a modified (weighted) Minkowski L1.

The study appeared in Can and Duygulu (2011) was motivated from a shape representation method. They used already segmented word images provided with the datasets. The proposed methodology comprises 3 main steps. First, binarization is applied by using a threshold which is computed as the mean intensity value of the gray-scale image. Next, a contour extraction is used for each connected component in the binarized word image. Finally, a sequence of lines is created which is used as the descriptor for matching. The matching score is computed by first finding the distances between the line descriptors and then, summing all distances over the complete word image.

The basic premise in the work of Fornés et al. (2011) is that each word image is treated as a shape which can be represented using two models, namely the Blurred Shape Model (BSM) and the Deformable BSM. First, at preprocessing stage, segmented text lines are normalized by applying skew angle removal and slant correction. In the case of BSM,

the descriptor represents a probability distribution of the object shape. In the case of Deformable BSM, every image is represented with two output descriptors, a vector which contains the BSM value of each focus (equidistant distributed points) and the position of each focus. The proposed matching technique lies upon the movement of focuses so that its own BSM is maximized. It is shown that using Euclidean distance in both BSM and deformable BSM methods outperforms the use of DTW.

Fernández et al. (2011) used Loci features (Glucksman, 1969) along eight directions. They are computed on the skeleton of each word image which is achieved after a document image binarization and word image segmentation step. The similarity is computed using the Euclidean and Cosine distance.

The study in Diesendruck et al. (2012b) and Diesendruck et al. (2012a) was focusing on building a search system for 1930-40 US Census data. The process starts with binarization, morphological thinning and Hough transform to locate table lines since the documents are in table format. Thus, the segmentation is based on table lines. Since, each cell contains one word, the method is word-based segmented. Then, a signature vector was composed the first 10 coefficients of cosine transform of upper, lower and transition profiles. Since the data set is quite large and the response time should be reasonable, hierarchical agglomerative clustering with complete linkage is used to cluster the signature vectors.

The problem of sequential KWS was addressed in Fernández-Mota et al. (2014). In sequential KWS the ordered sequence of indices is taken into account for finding similar instances of words in a book. They experimented with descriptors that relate to a single writer scenario (BSM, HOG, nrHOG) as well as descriptors that relate to a multiple writers scenario (attribute-based approach).

The work in Howe (2013) is based on a part-structured modeling which aims to minimize a deformation energy required to fit the query to the word image instance. The process is initiated with a binarization. Then, skeletonization is applied to produce connected components of a single-pixel width. The endpoints and junctions of the skeleton are used to build a tree. An energy minimization of a function that comprises a deformation energy and an observation energy term is finally addressed.

The method in Kovalchuk et al. (2014) is the winner of H-KWS 2014 competition. First, they binarize the image by global thresholding and connected components are computed. Then, pruning of connected components is followed based on heuristics that rely upon properties of connected components. Using a regular grid of fixed size, they compute HOG and LBP descriptors which result in a 250D vector. A max-pooling process is then applied to the descriptor. The matching is made with L_2 distance.

In Wang et al. (2014) the authors initiate the process by applying the preprocessing step presented in Wang et al. (2013). They use a graph representation model which is based on the skeleton of each word image. In this graph, vertices are the structural interest points and the strokes connecting them are the edges. The value of vertices corresponds to the Shape Context descriptor while the value of edges corresponds to the length of stroke. The computation of similarity between two word images concern the similarity of graphs for each connected component existing in the query and the test word image which is used to guide the DTW computation.

The work in Zagoris et al. (2014) is based on spatial information from word images. First, gradient vectors are calculated. Because of the sensitivity of gradient to noise, an Otsu

like threshold is applied at gradient vectors. At the remaining points gradient orientation is calculated. Next, a linear quantization of gradients orientation to a desirable amount of levels follows. The quantization step also controls the amount of the final local points. After quantization the corner points are characterized as initial keypoints (kP). The final points are the dominant kPs according to Shannon entropy at an area were the kP is the center. After the final points have been selected each area around kPs is divided to 9 subareas. For each of these 9 areas using a voting system based on the weighted distance of each point to kP, a 3-bin histogram is created. The combination of these 9 histograms to a 27-bin histogram results in the descriptor, the Document-Specific Local Features (DSLF). At matching stage first, a normalization is applied for each word image. Then instead of a brute force search, a Local Proximity Nearest Neighbour (LPNN) search is used by taking into account the mean distance of each pixel from the mean center. Finally, Euclidean distance is applied between the kPs and the results are presented in an ascending order. In Zagoris et al. (2015), an extension of this work is presented using Relevance Feedback strategies (CombSum, CombMin, Probabilistic model). It is reported that the optimal results are achieved from CombMin model.

The goal in Papandreou et al. (2016) is to study the zoning features. Binarization and deslanting are first applied in the query and candidate word images as pre-processing steps. The zoning features are extracted after cutting the query word in vertical zones based on its length and pixel distribution along the horizontal axis and adjusting these boundaries with the corresponding zones in the candidate word image using DTW. In the sequel, the word images are normalized and their features, which are based on pixel density, are extracted. Finally, the final distance is the product of Euclidean distance of the two word images with the distance provided from DTW.

In Mondal et al. (2016) the author introduces a new matching technique the Flexible Sequence Matching (FSM) algorithm for KWS task. At the preprocessing stage, the document images are first binarized by an adaptive technique (Gatos et al., 2006), after binarization along with border removal is applied to obtain proper text boundaries (Stamatopoulos et al., 2010) and then a segmentation stage follows that partitions the documents in lines or pieces of lines, up to words or parts of words, depending on the experiment to be conducted. At features extraction stage grayscale and binary image are used to extract two types of features, namely column-based features and Slit style HOG (SSHOG) (Terasawa and Tanaka, 2009). Eight Column-based features, are extracted from the binary image. Finally, at matching stage FSM is applied. FSM is similar in spirit to DTW but it is less sensitive to local variations in the spelling of words and to local degradations effects within the word image.

In a recent work Retsinas et al. (2016), three variances of the Projections of Oriented Gradient (POG) descriptor (Retsinas et al., 2015) studied in the framework of the KWS problem. The first variant, the global POG (gPOG) is slightly different from POG, for which the main differences are: (i) it keeps different number of coefficients from DCT and (ii) has 6 projections. The second variant k-segmented POG (lPOG), first segments the word image to k overlapping images and then calculates the POG descriptor to each of them. The third variant, fusion POG (fPOG), as the name implies is a fusion of gPOG and lPOG descriptors. Finally, the Euclidean distance is used to attain the matching score.

It should be noted that at the preprocessing step, binarization, skew correction and height normalization were applied.

Textline Segmentation Context

In the paper Kolcz et al. (2000), the approach is motivated by the success of dynamic-programming based techniques for KWS in speech applications even when very limited keyword models are present. It relies upon a line segmentation method to achieve distinct textlines for each document image. The ink-density histogram and its Fast Fourier Transform (FFT) spectrum is used to determine the textlines as well as the skew of the page. For each textline, they extract the upper and the lower profile as well as transitional Features (number of transactions between background and body in each column of pixels). Then they used DTW to address matching in the KWS pipeline. They used heuristics to reduce computational time which were similar to the ones used by the work in (Manmatha et al., 1996).

Terasawa and Tanaka (2009) deals with language independent KWS method. They have chosen a line-oriented approach because (i) word segmentation is impossible in some languages and (ii) it can retrieve a hyphenated word that spans two lines. For each textline image, a narrow sliding window is used for feature extraction. A variation of HOG features is used, namely, the SSHOG features. Compared to the original HOG, the SSHOG uses a narrow window and the computed gradient is signed. For feature matching, DTW is used.

In Wang et al. (2013) the authors use a coarse-to-fine strategy. They first remove the noise with a smoothing filter and they apply textline segmentation based on Hough transform. At coarse step, they apply a sliding window at the size of query word. They extract 4 textural features, namely, projection profile, upper and lower border and orientation distribution of skeleton pixels. Then they use DTW for the first three features and Chi-square metric for orientation distribution. With an empirical threshold they chose the most similar to the query regions, the regions of interest. The fine step is applied to the regions of interest. In the fine step, morphological and topological properties are used. Morphological properties calculated using the Shape Context descriptor on selected interest points (branch points, starting/ending points and high-curved points). Topological properties are obtained from a skeletonized representation. The information of the properties is the input to a weighted distance function for which Linear Discriminant Analysis is used to automatically get the optimal weights.

Segmentation-free Context

In this section, segmentation-free methods are presented. The methods of this section are divided into training-based and training-free.

Methods Based on Training

The approach in Choisy (2007) is made to deal with KWS of isolated words on mail envelopes. It is character segmentation-free which relies on the dynamic creation of global

word models. This is achieved with the use of Non-Symmetric Half Plane - HMM (NSHP-HMM) (Saon and Belaïd, 1997) which is a model hybrid of an HMM and a Markov Field (MRF). Before applying NSHP-HMM, two preprocessing steps are applied (i) global slant correction and (ii) a non-linear normalization that centers and normalizes the lower case zone of the writing. The NSHP-HMM is trained at word level.

In the work of Rusinol et al. (2011), they used a BoVW model where each patch is normalized by applying term frequency-inverse document frequency (tf-idf) model. BoVW words model powered by SIFT descriptors which was further augmented by a word segmentation. Then, Latent Semantic Indexing (LSI) (Deerwester et al., 1990) is used with BoVW to retrieve relevant patches even if they do not contain the same exact features than the query sub-image. Then Singular Value Decomposition (SVD) is applied to reduce the descriptor dimension and to obtain a transformed space where patches having similar topics but with different descriptors will lie close. At the retrieval stage cosine similarity were used. In Rusinol et al. (2015), they have enhanced their preliminary version by including an indexation scheme aimed to scale the proposed method to handle large datasets. The SVD step replaced by Product Quantization (PQ) for this purpose. Also, a multi-length patch representation is also introduced, which increases the retrieval performance by taking into account the different possible lengths of the query words.

The study presented in Rusinol and Lladós (2012) comprises both fusion and relevance feedback mechanisms. At fusion stage, three different fusion methods were tested, namely, early fusion, combMAX and Borda count. For relevance feedback, three different methods were tested, the Aocchio's algorithm (Cao et al., 2011), the Ide dec-hi method (Ide, 1971) and the relevance score algorithm (Giacinto and Roli, 2004).

The work in Almazán et al. (2012b) is based on exemplar SVM for better representation of features. Documents are represented with a grid of HOG descriptors, and a sliding window approach is used to locate the document regions that are most similar to the query. They use the Exemplar SVM framework to produce a better representation of the query in an unsupervised way. Finally, the document descriptors are compressed with Product Quantization (PQ) which has also the benefit of calculating the distance between the query and the quantized document with the use of a look-up table.

The method presented in Dovgalecs et al. (2013) may operate with words or graphics and is situated in the BoVW framework. In the offline stage, the BoVW is created from densely sampled SIFT features. In the online stage, candidate zones detection works by comparing query features with BoVW using chi-square distance. Then, the Longest Weighted Profile (LWP) algorithm which enforce spatial ordering information characteristics of words and graphical patterns alike, is used to compute the similarity score between the query image and the candidate zones.

The study in Rothacker et al. (2013) is based on the use of Bag-of-Features with HMMs. The method is divided into three parts. First, densely SIFT features are extracted in the whole document image and 5% of those are used to create a codebook size of 4096 for Bag-of-Features. Then, the Bag-of-Features representation feeds an HMM which encodes the sequential visual appearance of features that are located in the query bounding box. Finally, the document collection can be queried in a patch-based fashion where the output is a map of probabilistic scores from which the query results can be retrieved. As shown in the upper part of Figure 11.3 the document image representation is visualized. In the

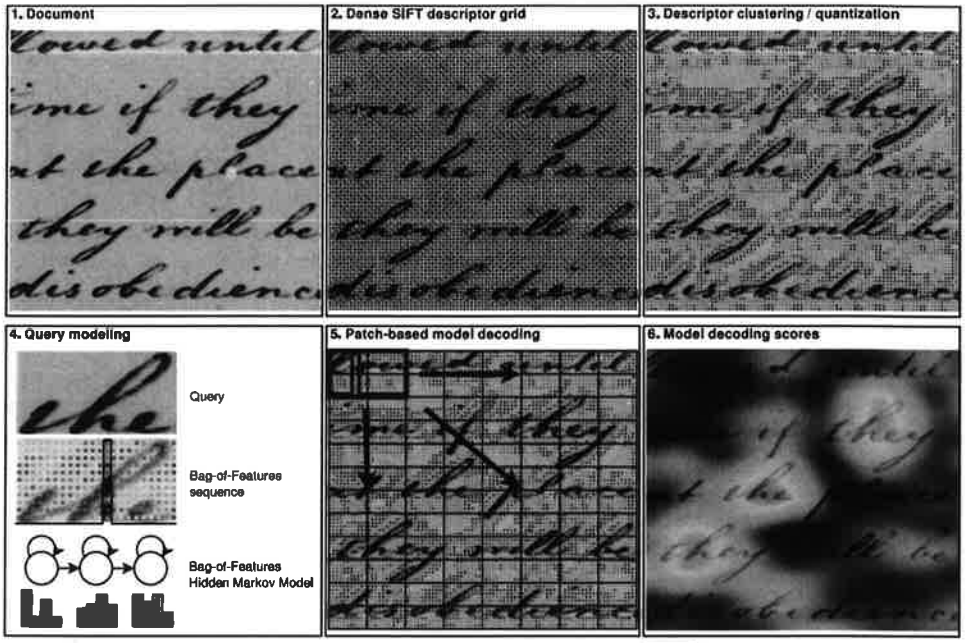


Figure 11.3. Overview of the segmentation-free word spotting method presented in (Rothacker et al., 2013).

lower part, the estimation of a query model and the patch-based decoding with respect to that model are shown. Patch-based representations are evaluated at each grid point. The scores obtained are visualized by interpolating them over the document image indicating low to high responses with blue to red colors.

Training - Free Methods

The work in Leydier et al. (2009) focuses on medieval Latin manuscripts and is based on the observation that medieval Latin manuscripts have letters mainly composed of large vertical strokes. The algorithm has two main steps. In the first step, guides, gradients and Zones Of Interest (ZOI)s are extracted from the document image and the query word image. In the second step, cohesive matching is applied between the guides of the query image and those of the document image. For each match of guides, a check if ZOIs are matching, is applied, too. This work was enriched with a model that is the combination of an alphabet, a glyph book and a grammar as presented in Leydier et al. (2009). The model is used to create a character tree. The extra information from the character tree was used among the query word image and the document image for better extraction of ZOIs, guides and gradients. They also automated thresholds that were needed at ZOIs, guides and gradients extraction, but also at cohesive matching.

Zhang and Tan (2013) is motivated by the Heat Kernel Signature (HKS) which has been used for shape recognition. Actually, the Deformation and Light Invariant (DaLI) descriptor

was used which is applied by convolving Scale Invariant HKS (SI-HKS) with Gaussian kernels. To compute the similarity, a Delaunay Triangulation algorithm was applied to create a Triangular Mesh Structure (TMS) of keypoints detected in the word image and the document image, respectively. Finally, the similarity score is computed by building a score matrix which contains the optimal matching score and the optimal matching history.

The approach in Hast and Fornés (2016) is based on Putative Match Analysis (PUMA) (Hast and Marchetti, 2012), a technique that first introduced for matching aerial images. First, the input images are binarized using Otsu, and then smoothed with a Gaussian in order to find more key points. Then, four different kind of key points are detected in the word images, which basically detect lines, corners and blobs. Taking into account the detected keypoints, Fourier-based feature descriptors are computed. At the end, the matching is performed by an improved version of Random Sample Consensus (RANSAC), called as PUMA, which is able to allow a more relaxed matching among the word images.

Rabaev et al. (2016) focus on KWS by locating the query word in a document recursively in a scale-space pyramid. The proposed scheme does not depend on a specific choice of features. They experimented with the HOG descriptors, which have been shown to provide good results. Chi-square distance is applied to compare HOG descriptors at all levels of the pyramid.

Experimental Datasets and Evaluation Metrics

Evaluation Metrics

The evaluation metrics that have been used for performance evaluation between different word spotting algorithms are inspired from information retrieval. Therefore, each retrieved item (word) is defined as relevant to original query (word - query) or not. Early reports on the KWS performance evaluation were simply taking the first n words and calculate the most basic retrieval metrics, the **Precision** and **Recall**.

$$Precision = \frac{\{relevant\ words\} \cap \{retrieved\ words\}}{\{retrieved\ words\}} \tag{11.1}$$

$$Recall = \frac{\{relevant\ words\} \cap \{retrieved\ words\}}{\{relevant\ words\}} \tag{11.2}$$

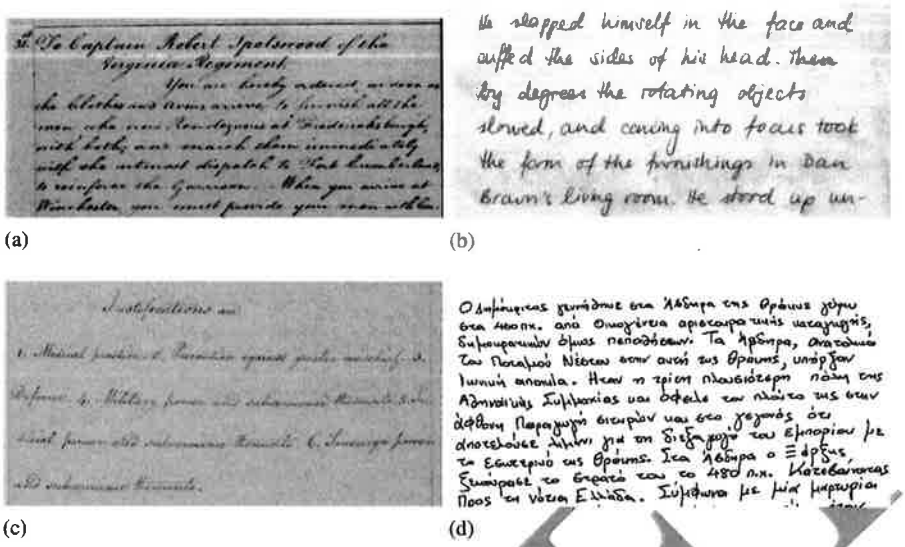
Precision is the fraction of retrieved words that are relevant to the search, while Recall is the fraction of the words that are relevant to the query. It is apparent that the above metrics are inversely related. To achieve a combined evaluation, the precision-recall curve is computed.

The **Precision - Recall Curve** is computed by the traditional 11-point interpolated average precision approach (Manning et al., 2008), (Van Rijsbergen, 1979). For each query, the interpolated precision is measured at the 11 recall levels of 0.0, 0.1, 0.2, ..., 1.0.

Sometimes, the differences between the evaluation algorithms are very hard to observe especially, between very small performance results. Moreover, these graphs may not

Table 11.1. Descriptors, learning methods and similarity measures used from each method

	Method	Descriptors	Learning	Similarity
Training based	(Rodríguez-Serrano and Perronnin, 2009)	LGH	SC-HMM	Euclidean
	(Rodríguez-Serrano and Perronnin, 2012)	LGH	SC-HMM	DTW
	(Almazán et al., 2012a)	Loci, HOG	SVM	Dot product
	(Fernández et al., 2013)	Loci	MLN	Euclidean
	(Aldavert et al., 2015)	HOG	BoVW	Histogram Matching
	(Sharma et al., 2015)	Deep features	CNN	Lp-norm
	(Keaton et al., 1997)	DCT on profiles	Bayesian network	Graph matching
	(Choisy, 2007)	Column-wise binary patterns	NSHP-HMM	Posteriori Probability
	(Rusinol et al., 2011)	SIFT	BoVW	Histogram Matching
	(Almazán et al., 2012b)	HOG	Exemplar SVM	Euclidean
	(Dovgalecs et al., 2013)	SIFT	BoVW	Chi-square
	(Rothacker et al., 2013)	SIFT	BoVW, HMM	Histogram Matching
Training free	(Manmatha et al., 1996)	Profiles		Euclidean, DTW
	(Adamek et al., 2007)	MCC, DCT		DTW
	(Bhardwaj et al., 2008)	Moments		Cosine
	(Zagoris et al., 2011)	CSPD		Minkowski L1
	(Can and Duygulu, 2011)	Sequence of lines		Line matching
	(Fornés et al., 2011)	Deformable		Euclidean, DTW
	(Fernández et al., 2011)			Euclidean, Cosine
	(Diesendruck et al., 2012b,a)	DCT on profiles		Euclidean
	(Howe, 2013)	Endpoints and junctions of skeleton		Energy minimization
	(Kovalchuk et al., 2014)	HOG, LBP		Euclidean
	(Wang et al., 2014)	SC		DTW
	(Zagoris et al., 2014)	DSLF		Euclidean
	(Papandreou et al., 2016)	Zoning features		Euclidean and DTW
	(Mondal et al., 2016)	Column-based, SSHOG		FSM
	(Retsinas et al., 2016)	POG, gPOG, lPOG, fPOG		Euclidean
	(Kolcz et al., 2000)	Profiles		DTW
	(Terasawa and Tanaka, 2009)	SSHOG		DTW
	(Wang et al., 2013)	Profiles, SC		DTW
	(Leydier et al., 2009)	ZOI		Cohesive matching
	(Zhang and Tan, 2013)	DaLI		Minimum cost path between connected keypoints in a mesh grid
	(Hast and Fornés, 2016)	Corners, blobs		PUMA
	(Rabaev et al., 2016)	HOG		Chi-square



(a) (b)

(c) (d)

Figure 11.4. Samples from the most used datasets (a) GW (b) IAM (c) Bentham (d) Modern

contain all the desired information (Salton, 1992). Therefore, the need to evaluate the retrieval results with a single value is needed. The most common evaluation metric that can meet this requirement is the **Mean Average Precision (MAP)** (NIST, 2013; Chatzichristofis et al., 2011) which is defined as the average of the precision value obtained after each relevant word is retrieved:

$$AP = \frac{\sum_{k=1}^{n}(P@K \times rel(k))}{\{relevant\ words\}} \tag{11.3}$$

where

Precision at k items (P@k) denoted as

$$P@k = \frac{\{relevant\ words\} \cap \{k\ retrieved\ words\}}{\{k\ retrieved\ words\}} \tag{11.4}$$

with the function of relevance denoted as follows:

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

Finally, the **Mean Average Precision (MAP)** is calculated by averaging the AP for all the queries, denoted as:

$$MAP = \sum_{q=1}^{Q} AP_q \tag{11.5}$$

where Q is the total number of queries.

It is worth to note that in the experiments for the segmentation-free case, a resulting 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:

$$OA = \frac{A \cap B}{A \cup B} \tag{11.6}$$

where OA is the overlapping area, A the resulting bounding box and B is the ground-truth.

The challenging nature of KWS in handwritten documents has motivated the organization of three dedicated international competitions in conjunction with the International Conference on Frontiers of Handwriting Recognition (ICFHR) and the International Conference on Document Analysis and Recognition (ICDAR). In particular, the ICFHR 2014 Handwritten Keyword Spotting Competition (ICFHR-2014) (Pratikakis et al., 2014), the ICDAR 2015 Competition on Keyword Spotting for Handwritten Documents (ICDAR-2015) (Puigcerver et al., 2015) and the ICFHR 2016 Handwritten Keyword Spotting Competition (ICFHR-2016) (Pratikakis et al., 2016) have been the venues where research groups have been competed in two different KWS scenarios, namely, segmentation-free and segmentation-based. Table 11.2, 11.3 and 11.4 shows the results for the ICFHR-2014, ICDAR-2015 and ICFHR-2016 competitions, respectively.

Table 11.2. Experimental results for the ICFHR-2014 Competition

	Segmentation-based				Segmentation-free			
	BENTHAM		MODERN		BENTHAM		MODERN	
Method	mAP	P@5	mAP	P@5	mAP	P@5	mAP	P@5
(Kovalchuk et al., 2014)	0.524	0.738	0.338	0.488	0.416	0.609	0.263	0.539
(Almazan et al., 2013)	0.513	0.724	0.323	0.706	-	-	-	-
(Howe, 2013)	0.462	0.718	0.2[illegible]	0.569	0.363	.556	0.163	0.417
(Leydier et al., 2009)	-	-	-	-	0.205	0.335	0.087	0.234
(Pantke et al., 2013)	-	-	-	-	0.337	0.543	0.091	0.245

Table 11.3. Experimental results for the ICDAR-2015 Competition

	Segmentation-based		Segmentation-free	
[illegible]	mAP	P@5	mAP	P@5
Pattern Recognition Group TU Dortmund University	.424	0.406	0.276	0.343
(Al[illegible] et al., 2013)	0.300	0.342	0.082	0.109

Table 11.4. Experimental results for the ICFHR-2016 Competition

	Segmentation-based		Segmentation-free	
	Botany	Konzil.	Botany	Konzil.
Method	mAP	mAP	mAP	mAP
Computer Vision Center (CVCDAG) Universitat Autoònoma de Barcelona, Spain	75.77	77.91	0	0
Pattern Recognition (PRG) TU Dortmund University, Germany	89.69	96.05	15.89	52.20
(Kovalchuk et al., 2014)	50.64	71.11	37.48	61.78
Visual Information and Interaction (QTOB) Uppsala University, Sweeden	54.95	82.15	-	-

Table 11.5. Segmentation type and Datasets used from each method.

	Method	Segmentation type		Dataset
		Line	Word	
Training based	(Rodríguez-Serrano and Perronnin, 2009), (Rodriguez-Serrano and Perronnin, 2012), (Rodríguez-Serrano et al., 2010)		X	French
	(Rodríguez-Serrano and Perronnin, 2012)		X	GW, French, IFN/ENIT
	(Almazán et al., 2012a), (Almazan et al., 2013)		X	CB
	(Fernández et al., 2013)		X	CB
	(Aldavert et al., 2015)		X	GW, Bentham, Modern
	(Sharma et al., 2015)		X	IAM
	(Keaton et al., 1997)	X		A
	(Choisy, 2007)	-	-	HMS
	(Rusinol et al., 2011, 2015)	-	-	GW
	(Almazán et al., 2012b)	-		GW
	(Dovgalecs et al., 2013)	-	-	GW
	(Rothacker et al., 2013)	-	-	GW
Training free	(Manmatha et al., 1996)		X	DIMUND, Hudson
	(Rath and Manmatha, 2003a,b, 2007)		X	GW
	(Adamek et al., 2007)		X	GW
	(Bhardwaj et al., 2008)		X	GW, IAM
	(Zagoris et al., 2011)		X	GW, Greek
	(Can and Duygulu, 2011)		X	GW, OTM
	(Fornés et al., 2011)		X	GW
	(Fernández et al., 2011), (Fernández-Mota et al., 2014)		X	CB
	(Diesendruck et al., 2012b,a)		X	USC
	(Howe, 2013)		X	GW
	(Kovalchuk et al., 2014)		X	GW
	(Wang et al., 2014)		X	GW, CB
	(Zagoris et al., 2014)		X	GW, Bentham, Modern
	(Papandreou et al., 2016)		X	GW, Bentham
	(Mondal et al., 2016)		X	GW, Japanese
	(Retsinas et al., 2016)		X	Bentham, Modern
	(Kolcz et al., 2000)	X		AIS
	(Terasawa and Tanaka, 2009)	X		GW, Japanese
	(Wang et al., 2013)	X		CITERE
	(Le[illegible] et al., 2007, 2009)	-	-	GW
	(Zhang and Tan, 2013)	-	-	GW
	(Hast and Fornés, 2016)	-	-	CB
	(Rabaev et al., 2016)	-	-	GW, CG, Arabic

"CB" stands for a collection of 50 pages from handwritten marriage licenses from the Barcelona Cathedral written in 1617'. URL: http://dag.cvc.uab.es/the-esposalles-database/

The "Bentham" dataset is part of the H-KWS 2014 contest's dataset. It consists of high quality (approximately 3000 pixels width and 4000 pixels height) handwritten manuscripts. The documents are written by Jeremy Bentham (1748-1832) himself as well as by Bentham's secretarial staff over a period of sixty years.

The "Modern" dataset is also part of the H-KWS 2014 contest's dataset. It consists of modern handwritten documents from the ICDAR 2009 Handwritten Segmentation Contest. These documents originate from several writers that were asked to copy a given text. They do not include any non-text elements (lines, drawings, etc.) and are written in four (4) languages (English, French, German and Greek).

A dataset that comprises 1539 pages of modern off-line handwritten English text written by 657 different writers is denoted as "IAM". URL: http://www.fki.inf.unibe.ch/databases /iam-handwriting-database

The Archives of the Indies in Seville (AIS), is a repository that represents the official communication between the Spanish Crown and its New World colonies and spans approximately four centuries (i.e. 15th-19th).

At Choisy (2007) a dataset that consists of French handwritten mails collection was used for Handwritten Mail Sorting (HMS) task, wherein 1522 mail pages are manually labelled.

At Manmatha et al. (1996) as dataset two single pages were used. One obtained from the DIMUND document server, thus denoted as "DIMUND" and the other single page was taken from a collection in the library of the University of Massachusetts. This page is a letter written by James S. Gibbons to Erasmus Darwin Hudson and it is denoted as "Hudson".

An Ottoman dataset denoted as "OTM" comprises documents written with a commonly encountered calligraphy style called Riqqa, which was used in official documents. Consists of 257 words in three pages of text. URL: http://courses.washington.edu/otap/

US Census forms from 1930 and 1940 comprise a dataset denoted as "USC".

Scanned images of the Japanese manuscript "The diary of Matsumae Kageyu" by Akoku Raishiki comprise a dataset denoted as "Japanese".

Letters written by different French philosophers constituting 4 collections are denoted as "CITERE". There are 11 pages containing approximately 2000 words, where 51 words were used as queries. URL: http://citere.hypotheses.org/

The Cairo Geniza (CG) collection that consists of 12 document images dated to the 10th century. This collection exhibits smeared characters, bleed through, and stains. The page size is about 1650×2330 pixels, and the collection contains 1371 words of 921 different transcriptions. URL: http://www.genizah.org/

A collection of 10 pages of Islamic manuscripts from Harvard University denoted as "Arabic" consists of documents that are dated from 12th to 15th centuries. The ground truth for this collection is given in terms of word-parts. Since word-parts are relatively small, for the experiments, 5117 largest word-parts were chosen with 929 different transcriptions. The page size is 1600 pixels. URL: http://ocp.hul.harvard.edu/ihp/

Conclusive Remarks

The major difference between segmentation-based and segmentation-free methods is the different search space (distinct word images versus the whole document image) where they operate. This is the basis of each disadvantage or advantage that each approach entails.

The main advantage of the segmentation-based methods is the retrieval speed. The ability to know the words boundaries inside the document provide profound advantages

with respect to an efficient retrieval performance. Therefore, segmentation-based methods are mainly based on word segmentations as there is only one method i.e. Keaton et al. (1997) which alternatively uses textline segmentation.

On the other hand, if the document is very noisy or very complex to apply a word segmentation method then a segmentation-free methodology is the most appropriate approach. Unfortunately, current segmentation-free KWS methods seem not to be appealing since the indexing storage size and the retrieval computation are very costly even when they are dealing with a medium size (100 pages) datasets. Moreover, the complicated issue of managing directly the whole document image is the main reason that few works deal with the segmentation-free approach.

Concerning features extraction, it is worth noting, that the majority of the works are relying on some form of gradients like SIFT, HOG, LGH, SSHOG, POG. Although, the initial approaches were using profile features, recent works understood the spatial texture information is more robust than shape especially for the handwritten documents. Lastly, some very recent works use features that are obtained using Deep Neural Networks which are called deep features.

For the training-based methods, the BoVW has been extensively used as a standalone learning component or combined with other models like the HMMs. The connection with the HMMs was motivated by the use of HMMs for handwritten transcription using a modelling inherent to the way a human makes a transcription. The recent advent of CNNs has started to appear in the KWS context (Sharma et al., 2015).

Until recently, the most common used dataset was the George Washington dataset for which, there was not a common evaluation protocol and each researcher employing its own subset and query set. Fortunately, the recent KWS competitions have set the ground for a concise performance evaluation framework.

Finally, in some works (Zagoris et al., 2011, 2014; Rusinol et al., 2011; Bhardwaj et al., 2008) the KWS pipeline is coupled with a relevance feedback mechanism which introduces the user in the retrieval loop, thus, improving the final retrieval performance.

Bibliography

Adamek, T., O'Connor, N. E., and Smeaton, A. F. (2007). Word matching using single closed contours for indexing handwritten historical documents. *International Journal of Document Analysis and Recognition (IJDAR)*, 9(2-4):153–165.

Aldavert, D., Rusinol, M., Toledo, R., and Lladós, J. (2013). Integrating visual and textual cues for query-by-string word spotting. In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pages 511–515. IEEE.

Aldavert, D., Rusiñol, M., Toledo, R., and Lladós, J. (2015). A study of bag-of-visual-words representations for handwritten keyword spotting. *International Journal on Document Analysis and Recognition (IJDAR)*, 18(3):223–234.

Almazán, J., Fernández, D., Fornés, A., Lladós, J., and Valveny, E. (2012a). A coarse-to-fine approach for handwritten word spotting in large scale historical documents collection. In *Frontiers in Handwriting Recognition (ICFHR), 2012 International Conference on*, pages 455–460. IEEE.

Almazán, J., Gordo, A., Fornés, A., and Valveny, E. (2012b). Efficient exemplar word spotting. In *Bmvc*, volume 1, page 3.

Almazan, J., Gordo, A., Fornés, A., and Valveny, E. (2013). Handwritten word spotting with corrected attributes. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1017–1024.

Bhardwaj, A., Jose, D., and Govindaraju, V. (2008). Script independent word spotting in multilingual documents. In *IJCNLP*, pages 48–54.

Can, E. F. and Duygulu, P. (2011). A line-based representation for matching words in historical manuscripts. *Pattern Recognition Letters*, 32(8):1126–1138.

Cao, H., Govindaraju, V., and Bhardwaj, A. (2011). Unconstrained handwritten document retrieval. *International Journal on Document Analysis and Recognition (IJDAR)*, 14(2):145–157.

Chatzichristofis, S. A., Zagoris, K., and Arampatzis, A. (2011). The trec files: the (ground) truth is out there. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 1289–1290. ACM.

Choisy, C. (2007). Dynamic handwritten keyword spotting based on the nshp-hmm. In *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on*, volume 1, pages 242–246. IEEE.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6):391.

Diesendruck, L., Marini, L., Kooper, R., Kejriwal, M., and McHenry, K. (2012a). Digitization and search: A non-traditional use of hpc. In *E-Science (e-Science), 2012 IEEE 8th International Conference on*, pages 1–6. IEEE.

Diesendruck, L., Marini, L., Kooper, R., Kejriwal, M., and McHenry, K. (2012b). A framework to access handwritten information within large digitized paper collections. In *E-Science (e-Science), 2012 IEEE 8th International Conference on*, pages 1–10. IEEE.

Dovgalecs, V., Burnett, A., Tranouez, P., Nicolas, S., and Heutte, L. (2013). Spot it! finding words and patterns in historical documents. In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pages 1039–1043. IEEE.

Fernández, D., Lladós, J., and Fornés, A. (2011). Handwritten word spotting in old manuscript images using a pseudo-structural descriptor organized in a hash structure. In *Iberian Conference on Pattern Recognition and Image Analysis*, pages 628–635. Springer.

Fernández, D., Marinai, S., Lladós, J., and Fornés, A. (2013). Contextual word spotting in historical manuscripts using markov logic networks. In *Proceedings of the 2nd International Workshop on Historical Document Imaging and Processing*, pages 36–43. ACM.

Fernández-Mota, D., Manmatha, R., Fornes, A., and Llados, J. (2014). Sequential word spotting in historical handwritten documents. In *Document Analysis Systems (DAS), 2014 11th IAPR International Workshop on*, pages 101–105. IEEE.

Fornés, A., Frinken, V., Fischer, A., Almazán, J., Jackson, G., and Bunke, H. (2011). A keyword spotting approach using blurred shape model-based descriptors. In *Proceedings of the 2011 workshop on historical document imaging and processing*, pages 83–90. ACM.

Gatos, B., Pratikakis, I., and Perantonis, S. J. (2006). Adaptive degraded document image binarization. *Pattern recognition*, 39(3):317–327.

Giacinto, G. and Roli, F. (2004). Instance-based relevance feedback for image retrieval. In *NIPS*, pages 489–496.

Glucksman, H. A. (1969). Classification of mixed-font alphabetics by characteristic loci. Technical report, DTIC Document.

Hast, A. and Fornés, A. (2016). A segmentation-free handwritten word spotting approach by relaxed feature matching. In *Document Analysis Systems (DAS), 2016 12th IAPR Workshop on*, pages 150–155. IEEE.

Hast, A. and Marchetti, A. (2012). Putative match analysis: a repeatable alternative to ransac for matching of aerial images. In *International Conference on Computer Vision Theory and Applications, VISAPP2012, Rome, Italy, 24-26 February, 2012*, pages 341–344. SciTePress.

Howe, N. R. (2013). Part-structured inkball models for one-shot handwritten word spotting. In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pages 582–586. IEEE.

Ide, E. (1971). New experiments in relevance feedback. *The SMART retrieval system*, pages 337–354.

Keaton, P., Greenspan, H., and Goodman, R. (1997). Keyword spotting for cursive document retrieval. In *Document Image Analysis, 1997.(DIA'97) Proceedings., Workshop on*, pages 74–81. IEEE.

Kolcz, A., Alspector, J., Augusteijn, M., Carlson, R., and Popescu, G. V. (2000). A line-oriented approach to word spotting in handwritten documents. *Pattern Analysis & Applications*, 3(2):153–168.

Kovalchuk, A., Wolf, L., and Dershowitz, N. (2014). A simple and fast word spotting method. In *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*, pages 3–8. IEEE.

Leydier, Y., Lebourgeois, F., and Emptoz, H. (2007). Text search for medieval manuscript images. *Pattern Recognition*, 40(12):3552–3567.

Leydier, Y., Ouji, A., LeBourgeois, F., and Emptoz, H. (2009). Towards an omnilingual word retrieval system for ancient manuscripts. *Pattern Recognition*, 42(9):2089–2105.

Manmatha, R., Han, C., and Riseman, E. M. (1996). Word spotting: A new approach to indexing handwriting. In *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR'96, 1996 IEEE Computer Society Conference on*, pages 631–637. IEEE.

Manning, C. D., Raghavan, P., Schütze, H., et al. (2008). *Introduction to information retrieval*, volume 1. Cambridge university press Cambridge.

Mondal, T., Ragot, N., Ramel, J.-Y., and Pal, U. (2016). Flexible sequence matching technique: An effective learning-free approach for word spotting. *Pattern Recognition*, 60:596–612.

NIST, T. (2013). Trec nist. *[Online]. Available: http://trec.nist.gov/pubs/trec16/appendices/measures.pdf*.

Pantke, W., Märgner, V., Fecker, D., Fingscheidt, T., Asi, A., Biller, O., El-Sana, J., Saabni, R., and Yehia, M. (2013). Hadara–a software system for semi-automatic processing of historical handwritten arabic documents. In *Archiving Conference*, volume 2013, pages 161–166. Society for Imaging Science and Technology.

Papandreou, A., Gatos, B., and Zagoris, K. (2016). An adaptive zoning technique for word spotting using dynamic time warping. In *Document Analysis Systems (DAS), 2016 12th IAPR Workshop on*, pages 387–392. IEEE.

Pratikakis, I., Zagoris, K., Gatos, B., Louloudis, G., and Stamatopoulos, N. (2014). Icfhr 2014 competition on handwritten keyword spotting (h-kws 2014). In *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*, pages 814–819. IEEE.

Pratikakis, I., Zagoris, K., Gatos, B., Puigcerver, J., Toselli, A. H., and Vidal, E. (2016). Icfhr2016 handwritten keyword spotting competition (h-kws 2016). In *Frontiers in Handwriting Recognition (ICFHR), 2016 15th International Conference on*, pages 613–618. IEEE.

Puigcerver, J., Toselli, A. H., and Vidal, E. (2015). Icdar2015 competition on keyword spotting for handwritten documents. In *Document Analysis and Recognition (ICDAR), 2015 13th International Conference on*, pages 1176–1180. IEEE.

Rabaev, I., Kedem, K., and El-Sana, J. (2016). Keyword retrieval using scale-space pyramid. In *Document Analysis Systems (DAS), 2016 12th IAPR Workshop on*, pages 144–149. IEEE.

Rath, T. M. and Manmatha, R. (2003a). Features for word spotting in historical manuscripts. In *Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on*, pages 218–222. IEEE.

Rath, T. M. and Manmatha, R. (2003b). Word image matching using dynamic time warping. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–II. IEEE.

Rath, T. M. and Manmatha, R. (2007). Word spotting for historical documents. *International Journal of Document Analysis and Recognition (IJDAR)*, 9(2-4):139–152.

Retsinas, G., Gatos, B., Stamatopoulos, N., and Louloudis, G. (2015). Isolated character recognition using projections of oriented gradients. In *Document Analysis and Recognition (ICDAR), 2015 13th International Conference on*, pages 336–340. IEEE.

Retsinas, G., Louloudis, G., Stamatopoulos, N., and Gatos, B. (2016). Keyword spotting in handwritten documents using projections of oriented gradients. In *Document Analysis Systems (DAS), 2016 12th IAPR Workshop on*, pages 411–416. IEEE.

Rodríguez-Serrano, J. A. and Perronnin, F. (2009). Handwritten word-spotting using hidden markov models and universal vocabularies. *Pattern Recognition*, 42(9):2106–2116.

Rodríguez-Serrano, J. A. and Perronnin, F. (2012). A model-based sequence similarity with application to handwritten word spotting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2108–2120.

Rodriguez-Serrano, J. A. and Perronnin, F. (2012). Synthesizing queries for handwritten word image retrieval. *Pattern Recognition*, 45(9):3270–3276.

Rodríguez-Serrano, J. A., Perronnin, F., Sánchez, G., and Lladós, J. (2010). Unsupervised writer adaptation of whole-word hmms with application to word-spotting. *Pattern Recognition Letters*, 31(8):742–749.

Rothacker, L., Rusinol, M., and Fink, G. A. (2013). Bag-of-features hmms for segmentation-free word spotting in handwritten documents. In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pages 1305–1309. IEEE.

Rusinol, M., Aldavert, D., Toledo, R., and Llados, J. (2011). Browsing heterogeneous document collections by a segmentation-free word spotting method. In *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, pages 63–67. IEEE.

Rusinol, M., Aldavert, D., Toledo, R., and Lladós, J. (2015). Efficient segmentation-free keyword spotting in historical document collections. *Pattern Recognition*, 48(2):545–555.

Rusinol, M. and Lladós, J. (2012). The role of the users in handwritten word spotting applications: query fusion and relevance feedback. In *Frontiers in Handwriting Recognition (ICFHR), 2012 International Conference on*, pages 55–60. IEEE.

Salton, G. (1992). The state of retrieval system evaluation. *Information processing & management*, 28(4):441–449.

Saon, G. and Belaïd, A. (1997). High performance unconstrained word recognition system combining hmms and markov random fields. *International Journal of Pattern Recognition and Artificial Intelligence*, 11(05):771–788.

Sharma, A. et al. (2015). Adapting off-the-shelf cnns for word spotting & recognition. In *Document Analysis and Recognition (ICDAR), 2015 13th International Conference on*, pages 986–990. IEEE.

Stamatopoulos, N., Gatos, B., and Georgiou, T. (2010). Page frame detection for double page document images. In *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*, pages 401–408. ACM.

Terasawa, K. and Tanaka, Y. (2009). Slit style hog feature for document image word spotting. In *Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on*, pages 116–120. IEEE.

Van Rijsbergen, C. (1979). Information retrieval 2nd edition butterworths.

Wang, P., Eglin, V., Garcia, C., Largeron, C., Lladós, J., and Fornés, A. (2014). A novel learning-free word spotting approach based on graph representation. In *Document Analysis Systems (DAS), 2014 11th IAPR International Workshop on*, pages 207–211. IEEE.

Wang, P., Eglin, V., Garcia, C., Largeron, C., and McKenna, A. (2013). A comprehensive representation model for handwriting dedicated to word spotting. In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pages 450–454. IEEE.

Zagoris, K., Ergina, K., and Papamarkos, N. (2011). Image retrieval systems based on compact shape descriptor and relevance feedback information. *Journal of Visual Communication and Image Representation*, 22(5):378–390.

Zagoris, K., Pratikakis, I., and Gatos, B. (2014). Segmentation-based historical handwritten word spotting using document-specific local features. In *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*, pages 9–14. IEEE.

Zagoris, K., Pratikakis, I., and Gatos, B. (2015). A framework for efficient transcription of historical documents using keyword spotting. In *Proceedings of the 3rd International Workshop on Historical Document Imaging and Processing*, pages 9–14. ACM.

Zhang, X. and Tan, C. L. (2013). Segmentation-free keyword spotting for handwritten documents based on heat kernel signature. In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pages 827–831. IEEE.