

A Brief Review of Document Image Retrieval Methods: Recent Advances

Fahimeh Alaei
School of ICT, Griffith
University, Australia
fahimeh.alaei@griffithuni.edu.au

Alireza Alaei
School of ICT, Griffith
University, Australia
alireza20alaei@yahoo.com

Michael Blumenstein
University of Technology
Sydney, Australia
Michael.Blumenstein@uts.edu.au

Umapada Pal
CVPR Unit, Indian
Statistical Institute, India
umapada@isical.ac.in

Abstract—Due to the rapid increase of different digitized documents, the development of a system to automatically retrieve document images from a large collection of structured and unstructured document images is in high demand. Many techniques have been developed to provide an efficient and effective way for retrieving and organizing these document images in the literature. This paper provides an overview of the methods which have been applied for document image retrieval over recent years. It has been found that from a textual perspective, more attention has been paid to the feature extraction methods without using OCR.

Keywords— *Document image retrieval; Document processing; Indexing; Similarity Matching.*

I. INTRODUCTION

Information for retrieval can be categorised into two different types: audio/speech and visual [1]. Visual data could be pictorial or textual, while images, graphs, diagrams, and maps are considered to be pictorial documents. In addition, textual data includes handwritten, printed, and complex documents [1]. Document image retrieval (DIR) is a research domain, which is marginal between classic information retrieval (IR) and content based image retrieval (CBIR) [2]. The task of document image retrieval is to find useful information or similar document images from a large dataset for a given user query. In this era, the trend has moved towards having a paperless world; hence, a significant number of documents, books, letters, historical manuscripts, and so on are saved through electronic devices in everyday life. These electronic images of paper-based documents are normally captured by scanners, fax machines, digital cameras, and mobile phones. The quantity of these data sets is dramatically increasing day-by-day. Automatic extraction, classification, clustering, and searching of information from such a large amount of data, is worthwhile. The last two decades have seen a growing trend towards document image retrieval to increase the efficiency, effectiveness, and speed of these methods. Still, finding a document from classified/unclassified data with an unconstrained structure is a crucial task. An overview of different techniques in the literature can be found in [1, 3, 4]. However, the purpose of this paper is to review the recent advances and research on textual and paper-based document retrieval.

Document image retrieval approaches are divided into two different groups: the recognition-based retrieval approach, which depends on the recognition of whole documents and the similarity between documents, is measured at the symbolic level; and recognition-free retrieval approaches [5-10], which rely on document image features, so that similarity is measured by the actual

content of the document images. Optical Character Recognition (OCR) is a traditional textual recognition method used for retrieval. The OCR-based approach has some weaknesses such as high computational cost, language dependency, and sensitivity to image resolution [11]. In the case of historical documents, which are usually of low quality, employing recognition-based approaches cannot provide appropriate results.

To deal with the drawbacks of OCR, each document image is represented as a feature vector for recognition-free retrieval. The same types of features are extracted for a query to complete the retrieval process. Therefore, retrieving similar documents to the query image without explicitly recognizing the documents is being attempted. Such a query design can be denoted as query-by-example, which has been computed at the raw data or feature level [11].

In Fig. 1, different steps, which have commonly been involved for document image retrieval in most of the methods presented in the literature, are demonstrated. The given block diagram shows two phases, training and testing. Firstly, pre-processing steps are provided to prepare suitable images for further analysis. Then, features are extracted at the coarse and fine levels; if dimension reduction is needed, appropriate methods should be applied in this step. The indexing/learning methods are applied to train a classifier or knowledge-based method for some given documents. Similarity distances between the query image and the documents in the dataset are measured, and finally the relevant image(s) matching the query image are displayed.

The rest of this paper is organized as follows. In Section 2, a variety of methods which have been applied for the pre-processing step in state-of-the-art methods for document image retrieval, are listed. Feature extraction, which is the most important part of retrieval, is discussed in 3. Section 4 is dedicated to the indexing and learning methods. Matching techniques and similarity distances applied in the last part of retrieval are considered in Section 5. A brief discussion on the results obtained in recent years is provided in Section 6, and finally conclusions are drawn in Section 7.

II. PRE-PROCESSING

Pre-processing is the first step of DIR. Since, document images may be noisy, distorted, and skewed, digitized documents need to be treated using different pre-processing methods. Pre-processing methods are divided into four main classes [12]: filtering, geometrical transformations, object boundary detection, and thinning.

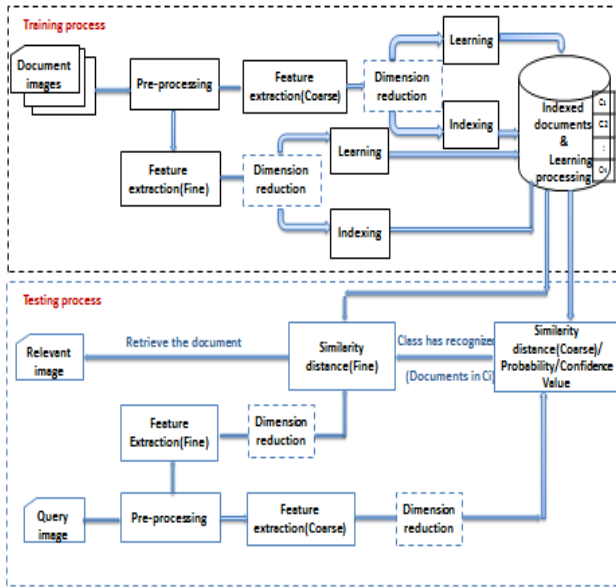


Fig.1. A general block diagram of document image retrieval.

According to the type of dataset, various pre-processing methods are applied to the document images. The filtering processes generally used in the literature are binarization, noise reduction, and signal enhancement [12]. Common noises in document images include excessive pepper and salt noise, large ink-blobs joining disjoint characters or components, vertical cuts due to folding of the paper and so on [13]. Mean filter [14] Median filter [15], and Gaussian filter [16] are the methods frequently applied to smooth document images. The smoothed images are commonly binarized by means of Otsu's or other algorithms [15, 17, 18]. Skew detection and correction [19-21], border removal [20], and normalization of the text line width [22] are also used to enhance document images. Moreover, in the initial steps, in some cases, colour images may be converted to grayscale images, and the sizes of images are reduced.

To find the skeleton of words for document image retrieval, thinning algorithms have been applied [15, 18]. These algorithms compute features based on the symbol skeleton and recursively erode the object contour.

III. FEATURE EXTRACTION

To enable an efficient search on document images, finding effective, unique and robust features is a crucial task. The extracted features significantly affect the retrieval performance [3]. Features used for document image retrieval are widely divided in two main categories: global features and local features.

A. Global features

Global features consider the whole document image for feature extraction. In other words, global features are visual features which can be further classified as general features and domain-specific features. In the case of document images, general features, such as texture, shape, size, and position of the document, have been considered for the retrieval process [23, 24].

The important information about the structural arrangement of each document and their relationships to the surrounding area are represented using texture features [25]. The visual texture properties are coarseness, contrast,

directionality, line likeness, regularity, and roughness. The wavelet transform is one of the methods for representing texture features. In [24], edge and texture orientations have been used as document image features. Also, multiscale and time-frequency localization of an image have been performed by wavelets. Since, the wavelets cannot represent the images with smooth contours in different directions, the Contourlet Transform (CT) method has been implemented by providing two additional properties, which are directionality and anisotropy.

Four types of texture features, namely multi-channel filtering features, fractal-based features, Markov random field parameters, and co-occurrence features, have been compared and evaluated in [26]. Some classification methods have been considered for assessment of the features. Co-occurrence features performed better in the given dataset as these resulted in a lower classification error [26].

Characterization of historical document images based on a texture feature has been presented in [27]. The extracted features were linked to the frequencies and orientations in different parts of a page. Physical or logical structures of the analysed documents were not taken into account in that study.

In [28], texture has been used to describe the types of features in document images, which have also become the search key for the document retrieval. Histogram of connected components and interest point densities over the documents have been used to compute texture features.

Shape representation-based features used for document image retrieval have been divided into two categories: boundary-based and region-based. For these two categories, the Fourier descriptor and moment invariants are, respectively, the most successful representatives, and are related by a simple linear transformation [29]. The finite element method (FEM) is another method that has been used for shape representation [30]. The FEM considers the connection of each point to other points on the object using a stiffness matrix. For the task of document image retrieval, shape representation as a visual feature is an important attribute. Shape context is computed for each point to describe the position of remaining points. The state-of-the-art shape representations, measures of shape dissimilarity, and shape matching algorithms have been discussed in [7].

To find the similarity between the layouts of documents, global features related to the position and sizes of a document with respect to other documents have been used in [5]. The extracted features have been saved in the feature vector and stored in a data-base management system (DBMS). In [31], the size and position of each block in a document have been defined, and then layouts have been considered for representing the class of each document using the Manhattan distance.

In [23], multi-scale run length histograms with the help of visual features have been considered as features for document image retrieval. The method is less sensitive to noise due to the use of visual features. In relation to the global features for document image retrieval, it can be noted that global features are robust, less sensitive to noise, and have good reliability. However, global features are less discriminative and they are not always unique.

B. Local features

Local features are extracted from a section of the document images. Depending on the document partitions, feature computation can be applied at different levels, for instance at the pixel level, column level, connected-component level, word, line, page level, and shape descriptor [3]. Since, feature extraction can be employed at different levels; the number of features varies case by case.

1) Pixel level features

By computing local features at the pixel level, some values will be dedicated to each pixel [27]. For the purpose of object detection, gradient descriptors have been used as a local feature. In the horizontal and vertical directions, the gradient of a two-variable function at each image pixel is a two-dimensional vector. Gradient-based binary features such as the gradient, structure, and concavity (GSC) have also been used in [32]. Each character image has been divided into 4×8 regions consisting of a 1024 bits (384-bits for gradient, 384-bits for structural, 256-bit for concavity) feature set. The correlation-based measure has been used for the similarity between two binary vectors. The authors of [32] claimed that retrieval using the GSC method provides faster and higher accuracy when compared to dynamic time warping (DTW), which uses profile-based features. In [33], word image retrieval has been performed using features such as the number of ink pixels in each column, location of the lowermost ink pixel, location of the uppermost ink pixel, and the number of ink to background transitions [32, 33].

Histogram of oriented gradient (HOG) is a technique which counts occurrences of the gradient orientation in the local part of an image. In [34], an extension of the HOG descriptor for a specific case of handwriting has been described. The combination of gradient features and a flexible plus adaptable grid has been used to extract features. Researchers have observed that better results were obtained for a word spotting method.

As a local feature at the pixel level, HOG features have been extracted in [35] for text retrieval. The potential characters have been detected with their location using HOG features extracted from sliding multi-scale window. A linear SVM classifier has been trained to spot characters of words in documents [35]. By using HOG features, explicit localization of the word boundary is not required to inform the document images.

2) Connected component-based features

In historical and handwritten documents, line and word segmentation are not easy tasks because of a variety of handwriting, touching, or broken characters [2]. Connected components-based features are important to deal with such document images. Commonly, after detecting the connected components of an image, based on the position and location of each connected component, further processes are also carried out. In the literature pertaining to DIR, many features were extracted based on the connected components of the images. In [36] word-spotting of old historical printed documents has been described and features, such as aspect ratio, horizontal frequency, number of branch points, scaled vertical centre of mass, height ratio to line height, and the presence of holes, have been extracted from the detected connected components.

In [37], hash tables have been built for indexing and compression using the connected component features of the document images. Component encoding in the hash table has been performed using components' contour points and a reduced number of interior points that are sufficient for component reconstruction.

In [38], text retrieval from early printed books carried out using character recognition is described. Characters have been recognized with connected component features as character objects. Occurrences of query words have been considered instead of recognizing the whole document. Self-organizing maps (SOM) have been used for data clustering, and then the similarity has been estimated with the help of the proximity of cluster centroids for retrieval purposes.

In [39], indexing techniques for text retrieval have been employed using connected component features at the coarse level. Approximate string matching algorithms have then been applied to find similar words in the document.

For each connected component as a character, width to height ratio, centre of gravity, horizontal/vertical projections, top-bottom shape projections, number of characters, top grid, and down grid features have been extracted in [15, 18]. The Euclidian distance method has calculated the distance between the query and the document images in the database for document retrieval.

In [11], a graph has been built for classifying document centroids of regions using connected component labelling and the centre of mass of all the regions. A Support Vector Machine (SVM) approach was applied to compute the probability that each document belongs to a specific class.

When considering connected component-based features, usually systems have high noise tolerance and less time consumption; however, degradation in historical documents can affect the results.

3) Word level features

In a local feature sequence and textual document image processing, words have a significant role for document image retrieval. To avoid the difficulties in character recognition and to enable faster approximation and computation, word level features have been applied for document retrieval. Usually, word level features are robust to image resolution but economical in terms of storage when a real-time retrieval speed is needed. However, features at this level do not produce intuitive results, and retrieval accuracy decreases when the size of the database is large. In addition, good results have not been obtained when font styles have dramatically changed. Words have usually been considered as a whole in word spotting applications. In [40], each word image has been represented by a fixed length sequence of vertical strips using word profile features. In [41], in addition to word profile features, height and width, baseline offset, and skew/slant angles have been extracted from word images. The features have then been normalized. In [14], the word length has been calculated by pixels and then the whole image has been represented as a single feature sequence instead of a big descriptor set. The centroid of each word region has been extracted as feature points [42, 64], and a locally likely

arrangement hashing (LLAH) feature vector has been calculated at each feature point. Word image matching for content-based retrieval has been proposed in [43]. The method is invariant to size, fonts and styles, and is suitable for printed documents.

In [44], the problems of font and style variation, where the query word image has a different style to the dataset, have been considered. A semi-supervised style transfer strategy has been proposed for reformulating the query word image using transfer learning.

4) *Zone level features*

Features can be extracted from a specific part of a page, through a fixed size window [45]. This technique has been used for supervised classification using a neural network. In [22], with the use of sliding window features such as moments of the black pixel distribution within the window, the positions of the black pixels, average grey scale and the number of vertical black/white transitions have been extracted for text lines. In [10], to capture the spatial relationship and correlation of the structure and layout of document objects, documents have been recursively partitioned based on image dimension, and speeded up robust features (SURF) have been extracted from each partition; then, documents have been encoded for classification and retrieval. SURF features that have been used at this level are scale invariant and robust to noise and distortion.

5) *Shape descriptors*

The scale-invariant feature transform (SIFT) has been applied in some previous research to characterize interesting points for document classification. In [46], after finding interest points, each descriptor has been indexed by its location in a uniform grid over the image. Descriptors have been clustered according to the index information. Then, matching of local features has been used to classify documents. In [47], word image retrieval has been performed using bag-of-visual-words. With the assistance of the SIFT method, salient points have been extracted and histograms of visual words have been created using hierarchical K-means clustering. The same features have been extracted in [48], and a pyramid histogram of oriented gradients (PHOG) has been created. The nearest neighbour classifier and the SVM method have been used for word image annotation.

A segmentation-free word spotting method using bag-of-features with a statistical sequence has been implemented in [49]. The SIFT descriptor has been applied to represent the documents, and each document page has been created by estimating a bag-of-features Hidden Markov Model (HMM).

Shape descriptors based on shape context have been implemented for document image indexing and retrieval in [9]. The Fourier-based shape descriptor has been introduced for the calculation of a hash index. The shape of an object in an image has been represented as a set of points. With the help of a logpolar histogram, relative arrangements of these points have been obtained and further used for document retrieval.

Signature-based document image retrieval has been presented in [7]. Shape context features have been computed for each point to describe the position of the remaining points. Subsequently, shape matching is

carried out while preserving the local neighbourhood structure for document image retrieval.

Shape descriptors are robust to size and are more reliable compared to pixel level analysis; also, in contrast, they are very sensitive to the results of segmentation and the type of writing.

With regard to features, local features are not always reliable but they are unique. Conversely, global features are reliable but not unique. Therefore, middle-level features can enable an appropriate trade-off [14].

IV. INDEXING/LEARNING METHODS

Automatic document indexing is an important issue in large collections used for document image analysis and retrieval. Classic indexing and retrieval can be divided into two parts: objective structured identifiers which consider titles, name, date, and publishers, and non-objective identifiers which can be extracted directly from the text content [4]. In addition, indexing a heterogeneous document can be through a physical or a logical structure.

Once documents are indexed, the resulting index vectors can be considered as signatures and used for retrieval [4]. In [38, 50, 51], indexing of words in old documents has been carried out using self-organizing maps (SOMs), and similar symbols have been clustered in a sub-set of the document.

In [61], classification of document images has been done based on visual similarity of layout structure. Type-independent features and geometric features have been extracted from document images. The decision tree classifier has been applied to provide semantically intuitive descriptions. Then, a neural network based SOM classifier has been used to find clusters in the input data as well as to detect each unknown datum with one of the clusters.

Neural network-based document image retrieval has been studied widely in [45, 62, 65, 66]. A layout-based document image retrieval system with the use of tree clustering based on an SOM neural network has been presented in [62]. Horizontal/vertical cuts along either spaces or lines have been considered as the internal nodes of the tree. Then, one vector-based tree representation has been used to train a SOM for clustering the pages on the basis of layout similarity. In [63], the SOM has been further considered for word clustering and word retrieval.

The classification capabilities of ANNs for layout analysis at pixel classification, region classification, and page classification have been compared in [45]. In [65], for identifying the complex document layouts, convolutional neural networks (CNN) have been applied. The CNN methods have been used to learn a hierarchy of feature detectors and train a nonlinear classifier. Document image classification and retrieval also have been carried out in the same way [66]. CNN approaches showed a better performance compared to BoW while larger datasets were available.

In [67], the words have been segmented and features have also been extracted using a time delay neural network (TDNN) to produce a segment membership score. The TDNN outputs have been used to form the membership matrix. Subsequently, dimension reduction has been employed to remove redundant bit vectors to

facilitate rapid nearest neighbour processing for indexing purposes.

For indexing the document images, shape descriptors based on shape context have been implemented and text and graphic regions in the document image have been identified [9]. Then, using horizontal/vertical projection profiles, text and word images have been segmented and for the calculation of a hash index Fourier-based shape descriptors have been applied. Similarly, in [37], a hash table has been created using connected components, which were extracted from shape features for document image indexing and compression. Component encoding in the hash table has been performed using component contour points and a reduced number of interior points.

SVMs have been applied for the retrieval process in [11, 35, 40]. For the most frequent queries, SVM classifiers have been used and a classifier synthesis strategy has been built for rare queries [40]. The one-shot learning scheme has been introduced to generate a novel classifier for rare/novel query words. In [35], by extracting HOG features, a linear SVM classifier has been trained. The characters of the words have been spotted and their score calculated based on the presence of the characters. An inverted index has been created which includes image identification and calculated score.

In the case of high variation and noise in datasets, SVMs cannot generalize well with sample training [10]; therefore other non-parametric methods can be used as classifiers.

V. SIMILARITY DISTANCE MATCHING

As previously explained, finding documents which are similar to a user query is the aim of the retrieval process. Similarity between query images and indexed document images can be performed at the pixel level or at the feature level. In both cases, the document image from the dataset that has a minimum distance with a query would be the most similar document image to the query image.

The nearest neighbour method has been commonly used to measure the similarity in some recent studies [40, 46, 48, 52, 53]. Euclidean and Manhattan distances have usually been applied to find distances between the feature vectors [5, 28, 48]. The Hamming distance [54] and Canberra distance [24] have also been considered to obtain similarity distances between the feature set of a given query and the feature sets of documents in a dataset. In [46], the nearest neighbour of each feature has been searched in a KD-tree and the similarity score for each document class has been computed by a number of nearest neighbour classifiers. Moreover, in [55] a segmentation method based on recognition has been employed and an approximate nearest neighbour search (ANNS) method has been considered for the feature matching phase.

The nearest neighbour-based segmentation algorithms have provided good results for the document with simple scripts and complex layouts. However, the results of documents with complex scripts and simple layouts are not satisfactory by using the nearest neighbour method because of the overlapping nature of the connected components [53]. The nearest neighbour classifier and the SVM method have been used for word image annotation

in [48], and the nearest neighbour method has provided more accurate results.

For retrieving word images using bag-of-visual-words (BoVW) [47], the scale-invariant feature transform (SIFT) method has been used to extract the features and to create the histograms. Then, Hierarchical K-Means (HKM) clustering has been applied for clustering of word images [47, 48].

In [56], the branch and bound search algorithm has been proposed for page classification through logical labelling graph matching. The tree edit distance computes the page similarity for layout-based document image retrieval in [57].

In [31], different block distances and matching methods have been compared and evaluated. Between the assignment problem, the minimum weight edge cover problem and the Earth Mover's distance, the minimum weight edge provided the better result.

In [17], a word shape coding technique has been presented for document image retrieval. By means of a vector space model, similarities between the query image and documents in the dataset have been computed using the cosine of the angle between vectors. In [19], for searching a query word, a sequence or a subsequence string of the query has been searched by inexact string matching. Then, similarities between a query word and word images extracted from the document have been measured based on dynamic programming to recognize the relevant word images. To deal with inexact matching, an additional term has been introduced to the formula in [50, 58], by considering the properties of the clustering algorithm.

VI. DISCUSSIONS

For an overview of the results of recent DIR methods in the literature, the results of recent studies are presented in Table I. From Table I, it can be noted that precision, recall, and F-measure have frequently been used in most of the papers as the evaluation metrics. Furthermore, only a few research groups have used some benchmarks, such as NIST, MARG, and Tobacco to evaluate their proposed DIR methods. Most of the research groups have, however, generated their own datasets for evaluating their proposed methods. Therefore, it is difficult to find a fair comparison study between the DIR methods proposed in the literature.

In relation to the type of features used for DIR, from Table I it can be noted that the global features provide better results in the case of complex and handwritten documents compared to the local features. This is because, important information about the structural arrangement of each document and their relationship can be obtained from global features. In addition, global features are robust to image resolution, image distortion and are language independent, so, these types of features can give promising results for the retrieval process. For printed books, which are usually structured documents, word level features and shape descriptors provide encouraging results. In text-to-image and camera-based document images, word level features also provided promising results; however, other feature levels, resulted in nearly 50% correct document image retrieval. Low accuracy has been obtained in historical documents. This

is mostly because historical documents are generally degraded and have poor quality.

Considering the system level analysis of the results obtained from different DIR methods in the literature, it can be noted that most of the methods are not scalable and suitable for a large volume of data. The document image retrieval methods relying on the layout analysis techniques can provide a powerful representation of document images; however, they are slow and error-prone for retrieval process [52, 57, 61, 62]. In addition, this type of method cannot distinguish the contents of document images and it is mostly appropriate for document images having structured layout. The document image retrieval methods [14, 31, 52, 57], which use variable-length descriptors and graph-based features for representation, have an effective and powerful feature representation. However, these methods are generally computationally expensive methods, as it is necessary to find distances between the variable-length descriptors such as sequences and graphs. The methods based on word spotting are faster compared to OCR-based approaches. They can further keep the structure of document images compared to the OCR-based approaches where the layout structure of the documents is distorted. They are also more suitable for large-scale and multi-lingual datasets. However, the segmentation

process to obtain appropriate word symbols is an error-prone process.

The neural network-based methods, which have an ability to implicitly detect complex nonlinear relationships between dependent and independent variables, can handle large amounts of data [65, 66]. The methods are computationally fast and not sensitive to input image size. So, neural network-based methods can provide promising results for document image retrieval [66].

VII. CONCLUSIONS

The purpose of this paper is to describe the recent work proposed for document image retrieval and to analyse their results. The focus was on recognition-free methods comparing the significantly different features which have been extracted at different levels for DIR, to provide better results in the retrieval process. Different steps that are generally involved for the document image retrieval process were also considered.

Further studies need to be carried out in order to develop/implement more intelligent and accurate approaches to gain easier and faster access to structured/unstructured data.

TABLE I. RESULTS COMPARISON OF RECENT PAPERS ON DIR

Data type	Features level	Dataset	Result	Method
Historical documents	Connected component-based features	Words from two different books of CESR collection	Up to 72% accuracy	[39]
		Göttingen, BLP, BLV, Munich	Top 10 up to 27%, 23%, 22%, 25% Res.	[38]
Printed books	Word level features	3 Scanned English books D1, D2, D3	Up to 98% accuracy	[40]
		5 Scanned books varying in font D1-D5	Up to 85% accuracy	[44]
	Shape descriptors	1000 Scanned books of Telugu	Up to 82.8% accuracy	[48]
		4 Different Indian languages	Up to 75.0% accuracy	[47]
Complex documents	Zone level features	Arabic document images, Tobacco litigation dataset, 5590 tax-form images	82% with 10 table images for training, 99% with 15 images for training	[10, 59]
		(CDIP) Tobacco Dataset	Text+ Image 80%	[60]
	Global features	NIST dataset MARG dataset	NIST baseline is 100% MARG baseline is 94.78%	[23]
	Connected component-based features	Girona Archives	Polar Graph with manual segmentation 98.81%	[11]
	Shape descriptors	NIST Special Database 2 and 6	For bank form database 99% and around 90% of the errors	[46]
Text-to-image and camera based	Pixel level features	SVT, ICDAR 2011 and IIIT scene text	mAP of 56.24% on SVT, 65.25% on ICDAR	[35]
	Connected component-based features	Various digital text documents	53.43% Mean Precision	[15, 18]
	Word level features	10 million pages	99.4% accuracy	[64]
Handwritten	Pixel level features	George Washington and Multi-language datasets D1, D2, D3	GW(50.19%), D1(89.27%), D2(82.04%), D3(80.72%)	[33]
	Global features	160 Specimen handwritten document	Top 1 is 100%	[24]

REFERENCES

- [1] Mitra, M. and B. Chaudhuri, Information retrieval from documents: A survey. Information retrieval, 2000. 2(2-3): p. 141-163.
- [2] Marinai, S., et al. A general system for the retrieval of document images from digital libraries. First International Workshop on Document Image Analysis for Libraries. 2004.
- [3] Marinai, S., B. Miotti, and G. Soda, Digital Libraries and Document Image Retrieval Techniques: A Survey. 2011. 375: p. 181-204.
- [4] Doermann, D., The Indexing and Retrieval of Document Images: A Survey. 1998: p. 40.
- [5] Cesarini, F., S. Marinai, and G. Soda, Retrieval by layout similarity of documents represented with MXY trees, in Document Analysis Systems V. 2002, Springer. p. 353-364.

- [6] Shin, C. and D. Doermann. Document Image Retrieval Based on Layout Structural Similarity. in IPCV. 2006. Citeseer.
- [7] Zhu, G., Y. Zheng, and D. Doermann. Signature-Based Document Image Retrieval. 2008: p. 752-765.
- [8] Tan, C.L., et al., Text retrieval from document images based on word shape analysis. *Applied Intelligence*, 2003. 18(3): p. 257-270.
- [9] Hassan, E., S. Chaudhury, and M. Gopal. Shape descriptor based document image indexing and symbol recognition. In 10th International Conference on Document Analysis and Recognition, 2009. ICDAR'09.. 2009.
- [10] Kumar, J., P. Ye, and D. Doermann. Structural similarity for document image classification and retrieval. *Pattern Recognition Letters*, 2014. 43: p. 119-126.
- [11] Gordo, A., et al. A kernel-based approach to document retrieval. In *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*. 2010.
- [12] Ha, T.M. and H. Bunke. Image processing methods for document image analysis. *Handbook of Character Recognition and Document Image Analysis*, 1997: p. 1-47.
- [13] Balasubramanian, A., M. Meshesha, and C.V. Jawahar. Retrieval from document image collections. in *Proceedings of the Document Analysis Systems*. 2006. Springer.
- [14] JilinLi, et al., Document Image Retrieval with Local Feature Sequences. *ICDAR*, 2009: p. 346-350.
- [15] Zagoris, K., K. Ergina, and N. Papamarkos, A Document Image Retrieval System. *Engineering Applications of Artificial Intelligence*, 2010. 23(6): p. 872-879.
- [16] Almazan, J., et al. A coarse-to-fine approach for handwritten word spotting in large scale historical documents collection. in *Proceedings of the International Conference on Frontiers in Handwriting Recognition (ICFHR)*, 2012.
- [17] Lu, S., L. Li, and C.L. Tan. Document image retrieval through word shape coding. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2008. 30(11): p. 1913-1918.
- [18] Zagoris, K., N. Papamarkos, and C. Chamzas. Web document image retrieval system based on word spotting. in *International Conference on Image Processing*. 2006.
- [19] Lu, Y. and C.L. Tan. Information retrieval in document image databases. *Knowledge and Data Engineering, IEEE Transactions*, 2004. 16(11): p. 1398-1410.
- [20] Gatos, B. and I. Pratikakis, Segmentation-free Word Spotting in Historical Printed Documents. 2009: p. 271-275.
- [21] Liu, H., et al. Document image retrieval based on density distribution feature and key block feature. in *Eighth International Conference on Document Analysis and Recognition*.. 2005.
- [22] Frinken, V., et al., A novel word spotting method based on recurrent neural networks. *Pattern Analysis and Machine Intelligence*, 2012. 34(2): p. 211-224.
- [23] Gordo, A., F. Perronnin, and E. Valveny, Large-scale document image retrieval and classification with runlength histograms and binary embeddings. *Pattern Recognition*, 2013. 46(7): p. 1898-1905.
- [24] Shirdhonkar, M. and M.B. Kokare. Handwritten Document Image Retrieval. in *Proceeding of 3rd International Conference on Computer Modeling and Simulation (ICCMS)*, Mumbai, India, pp. VI-506–VI-510. 2011.
- [25] Haralick, R.M., K. Shanmugam, and I.H. Dinstein, Textural features for image classification. *Systems, Man and Cybernetics*, 1973(6): p. 610-621.
- [26] Ohanian, P.P. and R.C. Dubes, Performance evaluation for four classes of textural features. *Pattern recognition*, 1992. 25(8): p. 819-833.
- [27] Journet, N., et al. A proposition of retrieval tools for historical document images libraries. in *Proceeding of Ninth International Conference on Document Analysis and Recognition. ICDAR*. 2007.
- [28] Cullen, J.E., J.J. Hull, and P.E. Hart. Document image database retrieval and browsing using texture analysis. in *Proceedings of the Fourth International Conference on Document Analysis and Recognition*.. 1997.
- [29] Rui, Y., T.S. Huang, and S.-F. Chang, Image retrieval: Current techniques, promising directions, and open issues. *Journal of visual communication and image representation*, 1999. 10(1): p. 39-62.
- [30] Pentland, A., R.W. Picard, and S. Sclaroff, Photobook: Content-based manipulation of image databases. *International journal of computer vision*, 1996. 18(3): p. 233-254.
- [31] Beusekom, J.V., et al. Distance measures for layout-based document image retrieval. In *Proceedings of the Second International Conference on Document Image Analysis for Libraries*, 2006. DIAL'06. Second International Conference on. 2006.
- [32] Zhang, B., S.N. Srihari, and C. Huang. Word image retrieval using binary features. In *Proceedings of the International Society for Optics and Photonics Electronic Imaging 2003*.
- [33] Nagendar, G. and C.V. Jawahar. Efficient Word Image Retrieval using Fast DTW Distance. In *Proceedings of the 13th International Conference on Document Analysis and Recognition (ICDAR)*, 2015.
- [34] Fernández-Mota, D., et al. Sequential Word Spotting in Historical Handwritten Documents. In *Proceedings of the 11th IAPR International Workshop on Document Analysis Systems (DAS)*, 2014 2014.
- [35] Mishra, A., K. Alahari, and C.V. Jawahar. Image retrieval using textual cues. In *Proceedings of the International Conference on Computer Vision (ICCV)*.. 2013.
- [36] Moghaddam, R.F. and M. Cheriet. Application of multi-level classifiers and clustering for automatic word spotting in historical document images. In *Proceedings of the 10th International Conference on Document Analysis and Recognition*.. ICDAR. 2009.
- [37] Chatbri, H. and K. Kameyama. Document image dataset indexing and compression using connected components clustering. In *Proceedings of the 14th IAPR International Conference on Machine Vision Applications (MVA)*.. 2015.
- [38] Marinai, S., Text retrieval from early printed books. *International Journal on Document Analysis and Recognition (IJDA)*, 2011. 14(2): p. 117-129.
- [39] Roy, P.P., F. Rayar, and J.-Y. Ramel. An efficient coarse-to-fine indexing technique for fast text retrieval in historical documents. In *Proceedings of the 10th IAPR International Workshop on Document Analysis Systems (DAS)*. 2012.
- [40] Ranjan, V., G. Harit, and C. Jawahar. Document retrieval with unlimited vocabulary. In *Proceedings of the Winter Conference on Applications of Computer Vision (WACV)*, 2015.
- [41] Rath, T.M. and R. Manmatha. Word image matching using dynamic time warping. In *Proceedings of the Computer Society Conference on Computer Vision and Pattern Recognition*. 2003.
- [42] Iwamura, M., T. Nakai, and K. Kise. Improvement of Retrieval Speed and Required Amount of Memory for Geometric Hashing by Combining Local Invariants. in *BMVC*. 2007.
- [43] Meshesha, M. and C.V. Jawahar, Matching word images for content-based retrieval from printed document images. *International Journal of Document Analysis and Recognition (IJDA)*, 2008. 11(1): p. 29-38.
- [44] Ranjan, V., G. Harit, and C. Jawahar. Enhancing Word Image Retrieval in Presence of Font Variations. In *Proceedings of the 22nd International Conference on Pattern Recognition (ICPR)*, 2014.
- [45] Marinai, S., M. Gori, and G. Soda, Artificial neural networks for document analysis and recognition. *Pattern Analysis and Machine Intelligence*, 2005. 27(1): p. 23-35.
- [46] Chen, S., et al. Structured document classification by matching local salient features. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR)*, 2012.
- [47] Shekhar, R. and C. Jawahar. Word image retrieval using bag of visual words. In *Proceedings of the 10th IAPR International Workshop on Document Analysis Systems (DAS)*, 2012.
- [48] Sankar, K.P., R. Manmatha, and C.V. Jawahar, Large scale document image retrieval by automatic word annotation. *International Journal on Document Analysis and Recognition (IJDA)*, 2013. 17(1): p. 1-17.
- [49] Rothacker, L., M. Rusinol, and G. Fink. Bag-of-features HMMs for segmentation-free word spotting in handwritten documents. In *Proceedings of the 21st International Conference on Document Analysis and Recognition (ICDAR)*, 2013.
- [50] Marinai, S., E. Marino, and G. Soda. Indexing and retrieval of words in old documents. In *Proceedings of the Seventh International Conference on Document Analysis and Recognition*, 2003.
- [51] Marinai, S., E. Marino, and G. Soda, Exploring digital libraries with document image retrieval. 2007.
- [52] Albert Gordo and Ernest Valveny, A Rotation Invariant Page Layout Descriptor for Document Classification and Retrieval. 2009: p. 481-485.

- [53] Kumar, K.S., S. Kumar, and C. Jawahar. On segmentation of documents in complex scripts. In Proceedings of the Document Analysis and Recognition. ICDAR. 2007.
- [54] Lu, S. and C.L. Tan. Keyword spotting and retrieval of document images captured by a digital camera. In Proceedings of the Ninth International Conference on Document Analysis and Recognition. ICDAR.. 2007.
- [55] Iwamura, M., M. Tsukada, and K. Kise. Automatic Labeling for Scene Text Database. 2013: p. 1365-1369.
- [56] Liang, J., et al. Page classification through logical labelling. In Proceedings of the 16th International Conference on Pattern Recognition. Proceedings.2002.
- [57] Marinai, S., E. Marino, and G. Soda. Layout based document image retrieval by means of XY tree reduction. In Proceedings of the Eighth International Conference on Document Analysis and Recognition. 2005.
- [58] Marinai, S., Emanuele Marino, and G. Soda, Word retrieval In Proceedings of the document images without OCR. 2003.
- [59] Doermann, D., P. Ye, and J. Kumar, Learning Document Structure for Retrieval and Classification. Conference on Pattern Recognition, 2012: p. 1558–1561.
- [60] Jain, R., D.W. Oard, and D. Doermann. Scalable ranked retrieval using document images. . In Proceedings of the International Society for Optics and Photonics IS&T/SPIE Electronic Imaging. 2013.
- [61] Christian Shin, D. Doermann, and A. Rosenfeld, Classification of document pages using structure-based features. International Journal on Document Analysis and Recognition, 2001. 3(4): p. 232-247.
- [62] Simone Marinai, Emanuele Marino, and G. Soda. Tree clustering for layout-based document image retrieval. In Proceedings of the Second International Conference on Document Image Analysis for Libraries. DIAL. 2006.
- [63] Simone Marinai, et al., Efficient word retrieval by means of SOM clustering and PCA, In Proceedings of the Document Analysis Systems VII. 2006. p. 336-347.
- [64] Kazutaka Takeda, Koichi Kise, and M. Iwamura, Real-Time Document Image Retrieval for a 10 Million Pages Database with a Memory Efficient and Stability Improved LLAH. 2011: p. 1054-1058.
- [65] Le Kang, et al. Convolutional Neural Networks for Document Image Classification. In Proceedings of the 22nd International Conference on. Pattern Recognition ICPR, 2014.
- [66] Adam W Harley, Alex Ufkes, and K.G. Derpanis, Evaluation of Deep Convolutional Nets for Document Image Classification and Retrieval.1502.07058, 2015.
- [67] Chellapilla, K. and J. Platt. Redundant bit vectors for robust indexing and retrieval of electronic ink. In Proceedings of the Ninth International Conference on Document Analysis and Recognition. ICDAR. 2007.