Word Spotting in Handwritten Document Images based on Multiple Features

Mallikarjun Hangarge, Veershetty C.

Abstract: ***This paper presents word spotting in handwritten documents based on multiple features. Multiple features are derived using Gabor, Histogram oriented gradient (HOG), Local binary pattern, texture filters and Morphological filters. The real time documents are heterogeneous in nature, for instance application forms, postal cards, railway reservations forms etc. includes handwritten and printed text with different scripts. To spot a word in such documents and retrieving them from a huge digitized repository is a challenging task. To address such issues word spotting based on multiple features is carried out with learning and without learning methods. In both the methods (learning and learning free) texture filters are exhibiting outstanding performance in terms of precision recall and f-measures. To confirm the capability of the proposed method, extensive experiments are made on publically available dataset i.e.GW20 and noted encouraging results compared to other contemporary works.***

Keywords: Document Image Processing, Image Retrieval, Cosine Distance, Optical Character Reorganization, Word Spotting

I. INTRODUCTION

A document is a very important source of information or evidence; in the worldwide, many physical documents are available offline which are less accessible to the end users. The age of the document, poor quality and its importance on historical aspect are the few reasons. Therefore, it is the responsibility of the digital libraries and other institutions to make these documents available online free for the end users. In this context, the present days, digital libraries and other institutes are working on this. They have been preserving huge documents in the form of an image without indexing or adequate information. To access or retrieve these documents from the database or information in the documents, we need metadata of the database. In this regard, robust indexing algorithms are necessary for fetching the relevant documents from the huge storage of the digital libraries. To index these documents, there are two methods: one is Optical Character Recognition (OCR) and second is word spotting. Basically, OCR works at the character level and transcripts the complete document for editing, indexing, searching etc. OCR results are poor in case the documents are of poor quality, complex in layout and multi-font and multi-sizes. Moreover, it is not suitable for the Indian documents because they are very complex in nature as they include complex layouts, multi-scripts, multi-fonts, etc. On the other hand, word spotting is an alternative to OCR which facilitates document retrieval without transcription. It is robust in retrieving and indexing the documents compared to OCR. This method works on searing and marching the query words to retrieve the documents. The concept of word spotting was firstly presented by Manmatha et al. [1]. Later on, several words matching procedures were published [2, 3]. Meanwhile, there is a number of issues still exists with the word spotting algorithms, though it is superior to OCR in document retrieval. Some of the major issues with word spotting algorithms where they have poor performance are document quality, type of documents such as handwritten, printed, multi-script, multi-font documents etc. Besides, these algorithms will not work efficiently in case of large-scale archives database and heterogeneous database. They fail to provide information about the number of documents indexed per second, how many queries are given and retrieved with respect to time, update rate, information retained etc...

In this paper, we have made a few interesting empirical studies that are carried out in different scenarios. Most of the reported works on word spotting were attempted on a single scenario such as spotting words in printed or handwritten documents independently. But, in the real situation, the digitization of the documents and types of documents taken for digitization are hybrid in nature. The process of capturing documents may happen with scanner or camera or mixed. The documents under digitization may be of printed or handwritten or mixer of these. This real situation creates a heterogeneous database. So, retrieving the words from such a database is certainly a challenging problem. In this context, we have made exhaustive experiments in different scenarios. Similarity matching and learning based experiments are carried out with different feature extraction methods on public and own datasets. Especially, the camera captured documents are used for experimentation to observe the performance of the proposed methods. Mainly, the proposed method is based on zoning of the image which leads to capture the minute information required for discrimination of patterns. Further, global and local properties of the image are used independently and in combination to describe the query and candidate words. On each zone, we computed HOG, LBP, Gabor, Morphological and Texture properties.

This paper is organized as follows. In Section 2, we have summarized previous work. Then, we presented the methods in detail in Section 3.

Revised Manuscript Received on October 30, 2019.
***** Correspondence Author

Mallikarjun Hangarge*, Department of Computer Science, Karanata Arts, Science and Commerece College, Bidar, India, mhangarge@yahoo.co.in

Veershetty C., Department of Computer Science, Government First Grade College, Basavakalayan, Bidar, India, vshetty1180@gmail.com

Experimental results are presented in Section 4. Finally, Section V contains the comparative analysis and Section 6 conclusions respectively.

II. RELATED WORK

Seung-Ho Lee et al. [4] has proposed a technique based on island driven lattice search system for online English cursive script. In this method, HMM is used to represent the geometric feature of the letters to spot on the document images, the spotting process were accomplished by the Viterbi search algorithm. The experiment was performed on a dataset of 22900 words and obtained the result at 85.4% accuracy. Patricia Keaton et al. [5] worked on automatic handwritten documents retrieval based on local and global descriptors such as cavities and profile signatures. The database used for experimentation consists of multi-writers handwritten pages. The overall classification accuracy reported is 78%. Another method for word spotting in handwritten documents is presented by A. Kołcz et al. [6] based on shape features. T. M. Rath et al. [7] have worked on handwritten documents indexing. To do this, they have used George Washington's Manuscripts database. The indexing is in terms of an average precision rate is 73%. Further, Christophe Choisy [8] has proposed a method for a keyword spotting based on the NSHP-HMM. The French handwritten emails of 1522 pages are used for experimentation. These pages are labeled manually. The testing set consists of 508 pages which yield 22738 segmented words and remaining pages are considered for training purpose. The method showed good results in terms of average precision. Handwritten word image retrieval method was presented by Jose A. Rodrıguez-Serrano et al. [9]. In this method, they used the synthesized query to retrieve similar words from document images. The features employed are (i) local gradient histogram (LGH) features, (ii) semi-continuous hidden Markov models (SC-HMM). The experiment results were presented in terms of mAP rate at 64.0%.The word spotting on Grayscale Pashto documents written in Arabic script was presented by Muhammad Ismail Shah et al. [10]. They used profiles and transitional features of each word. The correlation similarity is used to match the query word with all other words of the database. The experiment results are presented in terms of average precision rate as 94.75 % and average recall as 60.25%. Word spotting in chines handwritten documents is proposed by in the year 2013 by Liang Huang et al. [11]. The method used relative word models which compute the similarity between the words and models. The models are represented by geometric settings and linguistic settings. The geometric setting model describes the single-character similarity and between-character relationship. The linguistic model gives the dependency of the word with the outside adjacent characters. Experiments were carried out on a huge database which validates the effectiveness of the proposed experiment and justifies the benefits of geometric and linguistic settings. As compared to transcription-based text search, this method gives higher precision and recall rate. Marcal Rusinol et al. [12] segmentation free word spotting method is proposed based on four different handwritten and printed historical documents. They used a patch-based context where the local patches are defined by bag-of-visual-words. The experimental results on both handwritten and typewritten documents are encouraging. Leonard Rothacker et al. [13] have proposed a word spotting algorithm which allows exploring the document images without full transcription. The partial word matching is also obtained by considering the context of the query by string method. They utilized learn context-dependent character models from the training set that was very small with respect to the number of word models. This is feasible due to the use of Bag-of-features and HMMs that are particularly suitable for approximating the robust models with the only limited training set. In the opposite query-by string methods they consider a fully segmentation-free decoding model that does not need any pre-segmentation of the word or line. Experiments were performed on George Washington benchmark dataset. Results reported are in terms of average precision rate. Suman K [14] has proposed a method for segmentation-free word spotting based on convolution neural network. They applied a simple pre-segmentation method to yield a set of word instances which are then sent through a convolution neural network to predict PHOC embedding for all instances in a single forward pass in the network. Word matching is performed by the use of the nearest neighbor method. And they have achieved good results. Document indexing and retrieval method were proposed by Hafiz Adnan Niaz et al. [15].This method uses DCT and DWT features. A total of 30 English documents which covers more than 20,000 strings are considered. To evaluate the performance of the proposed method, 100 query words are given to the proposed system and obtain a recall rate as 87% and accuracy rate as 93%. From the above literature review, we can understand that the major word spotting and document retrieval works were carried out with English scripts. A little work reported on the rest of the global languages such as Chinese and Arabic. Particularly in Indian context little research work can be found with other than south Indian scripts. But there is no work carried out on south Indian documents particularly for Kannada script. Therefore, we motivated and proposed a method which presents an effective way for word spotting in Kannada handwritten documents using zone-based textural features.

III. PROPOSED METHOD

In this section, the task of word retrieval is carried out in two different scenarios one is without learning and another is with learning. Fig.1 shows the different steps involved in learning free word retrieval. Here, the document image is binarized using Otsu's algorithm which was used in [16]. Afterward, each word of a document image made as a single connected component by filling the gap between characters. This is done using directional dilation. Bounding boxes are fixed on each word and then extracted these connected components. In this way, 100% segmentation is achieved. However, small components such as commas, semicolon, single and two character words are discarded based on their aspect ratios [17]. Further, these word images are divided into two levels equally called patches (20 patches for one-word image), after this, we considered a patch of the word image and extracted various features independently. Further, patch features are combined to yield a single feature vector which represents the whole word image.

The feature vectors corresponding to the query word and the word in the database are matched using a cosine distance. The similarity scores are stored in a similarity matrix. The similarity matrix is used to locate the presence of the query words in the corpus.

Finally, completely and partially matched words are retrieved, and unrelated words are filtered out by on distance thresholding. Most of the following feature extraction methods have been used by the number of researchers. However, zoning and the combination of these features are outperforming in word and document retrieval based on learning and learning the free. Besides, we applied these methods to the heterogeneous dataset. Particularly, Kannada Multi-writer documents are used. Our aim is to study the performance of the multiple methods and their combination on public and private datasets with different scripts. The process of spotting the word and retrieving related document images becomes more complex when a document contains different scripts, printed and as well as handwritten text, graphical components etc. In such cases these, methods are not robust.

1. Gabor features
2. Histogram oriented gradient(HOG)
3. Local binary pattern features
4. Morphological features
5. Combination of 1 and 4
6. Combination of 2 and 4
7. Combination of 3 and 4 and lastly
8. Texture filters

In learning method, we have divided the segmented word database into two classes one is query class and another is non-query class. The feature extraction method remains the same. Using medium tree and Gaussian support vector machine classifiers the word retrieval task has been accomplished. Fig. 1 and Fig. 2 show the pipeline of the learning free and learning-based word retrieval method

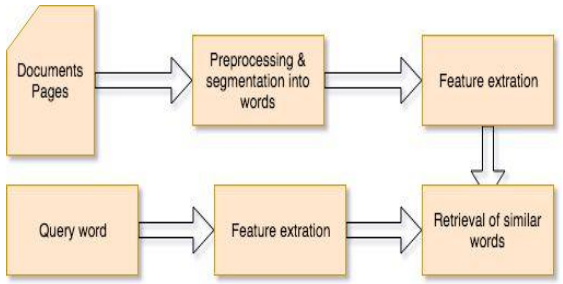

Documents Pages
Preprocessing & segmentation into words
Feature extration
Query word
Feature extration
Retrieval of similar words

Fig. 1.Different stages of the traditional word retrieval method

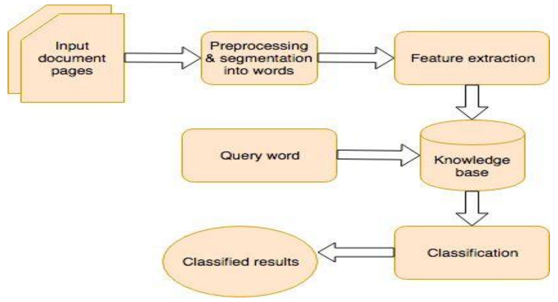


Fig.2. Different stages of learning based word retrieval method

IV. Experiments

4.1 Dataset

For the purpose of experimentation, we have considered four datasets and they are named as A, B, C and D. A- George Washington (GW) dataset which is publicly available [18]. B- Multi-writer handwritten Kannada dataset which is created by us. C- Camera captured handwritten Kannada documents. D- Heterogeneous dataset i.e. mixture of all these.

4.1.1 Dataset A

The George Washington database is created from the pages collected from the Library of Congress. These documents belong to the 18th century and written in English script, the size of the database is 20 pages. And these documents are written by two writes with a longhand script. The segmented words of this dataset are used for experimentation. Fig. 3 shows segmented few lines and words of George Washington dataset.

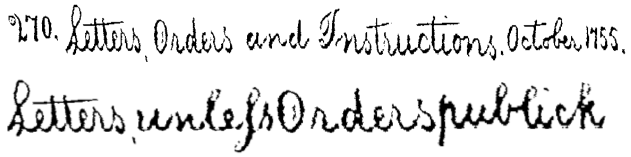

270. Letters Orders and Instructions October 1755.
Letters unless Orders publick

Fig.3 (a) sample line image of the GW database (b) segmented word of the GW database

4.1.2 Dataset B

This dataset is created using 150 handwritten pages. These pages are written by 50 writers of different ages (every 3 pages) with different pens. These documents are scanned at 300 dpi resolution from HP scanner. All pages are binarized using Otsu's algorithm. Morphological filters are used (opening) for noise removal. Resulting document images are considered for word segmentation and it is carried out as explained in Section 3. In this way generated 13000 words. All these isolated words are stored in the library with adequate indices. The Fig. 4(a), Fig. 4(b) and Fig4 (c) illustrate a sample document page, a binary image with the word as a single connected component and segmented word images respectively.

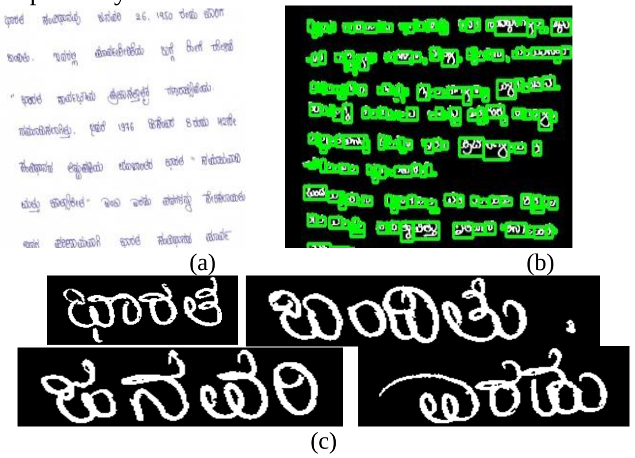
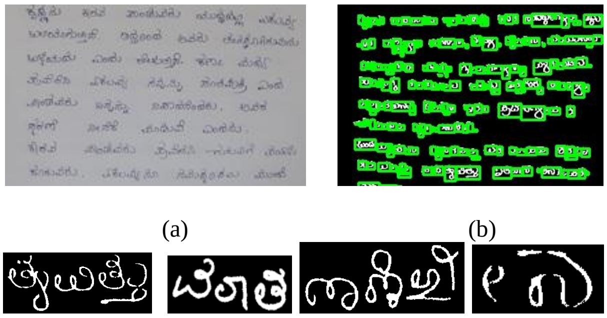


Fig. 4(a) Sample document page (b) Binary image with connected component and bounding box, (c) Example of segmented word

4.1.3 Dataset C

It is a camera captured dataset. Taken 50 multi-writer handwritten pages and captured using cell phone camera (Samsung J16 series). Obtained 4700 words from these documents after segmentation as explained in section 3. All these isolated words are stored in the library with adequate index. The Fig. 5(a), Fig. 5(b) and Fig 5(c) show a sample document page, a binarized image with bounding boxes on each word and segmented word images respectively. Table 1, summarizes the details about the datasets used in this paper



(a) (b)

(C)

Fig. 5(a) Sample document image (b) a binarized image with bounding boxes on each word (c) Segmented words

4.1.4 Dataset D

Dataset D is heterogeneous (the mixture of printed and handwritten) in nature. In digitization process of real world documents we come across with various types of documents such as application forms, postal cards, railway reservation forms, land records etc., where both handwritten and printed text exist. Moreover, cell phone is a portable device, so digitalization of such documents can be done easily through it to deal with real situations where the scanner cannot be carried out. To consider all these situations, this dataset has been created. This dataset is created using 25 printed pages. These pages are collected from the library with different books of the regional language that is Kannada. By segmenting the digitized pages generated 5000 words. All these isolated words are stored in the library with adequate indices. Along with this, we added segmented words of the previous datasets B and C, to make the dataset heterogeneous in nature. Fig.5 illustrate the samplewords of the hetrogeniuos dataset.

Fig. 5 illustrate the samplewords of the hetrogeniuos dataset.

Table-I. Summary of the databases.

Database	Pages	writers	century
Database A	20	1-2	18th
Database B	150	50	20th
Database C	50	>5	20th
Database D	25	printed	20th

4.2 Feature extraction methods

Feature extraction is the most important step of any automatic pattern recognition/matching system. Several feature extraction algorithms have been proposed by the number of researchers for document image analysis. These features can be broadly classified into local and global features. Local features describe the patches of the underlying image whereas the global features describe the image as a whole with a single vector. Multiple points of the image are used to compute the local features. In document image processing, both the types of features have been used. In most of the algorithms, the combinations of these features have been proposed. In majority of the reported works the combined features yielded notable results [19, 20]. Consequently, the probability of getting interesting results with different possible combinations is high. In this context, we have employed five different conventional global and local methods with combination and without combination (Independent) on different datasets. All the methods and their combinations are explained in the following Sections.

4.2.1 Gabor wavelets

In 1946, Dennis Gabor has presented a Gabor function [21].It is one of the efficient methods for texture and biometrics analysis. These Gabor wavelets give the illustration of the whole image rather than its parts in terms of pixels [22]. In a 2D-surface the total square of a correlation between an image and the two-dimensional Gabor function gives native spectral energy density associated with the assumed point and frequency of the image. There are numerous methods are available for using Gabor functions or wavelets. Two common methods are edge detection or curve finding, by a combination of responses to certain filters with a different position of the image. The important property of the wavelet is that it reduces the production of its standard deviations with respect to time and frequency. Gabor filters (or Gabor wavelet) generally applied to extract energy features from the images for image retrieval problem and has been shown to be very effective. Manjunath et.al [31] have shown that image retrieval using Gabor features are robust as compared to the pyramid-structured wavelet transform (PWT) features, tree-structured wavelet transforms (TWT) features and multi-resolution simultaneous autoregressive model (MRSAR) features. Gabor filters with its two-dimensional special function and its two dimensional Fourier transforms mathematical representation as follows

$$G(a,b,F,\sigma_a,\sigma_b) = 1/(2\pi\sigma_a,\sigma_b)\exp[-\frac{1}{2\left(\frac{a^2}{\sigma_a^2}+\frac{b^2}{\sigma_b^2}\right)} + 2\pi jF_a], \tag{1}$$

$$\mathrm{G}(u,v) = exp\{-\frac{\frac{1}{2[(u-F)^2}}{\sigma_a^2} + \frac{v^2}{\sigma_v^2}\} \tag{2}$$

where, $\sigma_u = \frac{1}{2\pi\sigma_a}$, $\sigma_v = \frac{1}{2\pi\sigma_b}$ and parameter F specifies the central frequency of interest. Gabor wavelets are derived from equation (1) as defined

$$g_{mn}(a,b) = x^{-m} g(a'b', \frac{Fn\pi}{k}, \sigma_a \sigma_b) \tag{3}$$

where $a' = x^{-m}(acos\emptyset + bsin\emptyset)$,
$b' = x^{-m}(asin\emptyset + bcos\emptyset)$.

In equation (3), m = 0, 1…; S-1, where S is the total number of scales, and n = 0, 1…; K -1, where K is the total number of orientations. The central frequency of interest is F. Then Gabor wavelets is calculated on each word image using equation (3) with 4 different scales and 6 orientations which yields the 24 feature set for each word of the document image and therefore, an angle of Orientation i.e., $\emptyset$(theta) is defined below:

$$\text{Orientation } (\emptyset) = (\frac{n\pi}{k}) \tag{4}$$

The parameters used in the above equations are defined as follows:

$$x = (U_h/U_l)^{\frac{1}{s-1}}$$

$$\sigma_{a=}((x+1)\sqrt{}2\ln 2)/2\pi x^m(x-1)U_l$$

$$\sigma_{b=}1/(\frac{2\pi\tan\left(\frac{\pi}{2k}\right)\sqrt{U_h^2}}{2\ln 2} - \left(\frac{1}{2\pi\sigma_a}\right)^2$$

By changing m and n, we have employed filter $g_{mn}(a,b)$ to the input image I (using 2-D convolution) to get features at different scales and orientations. In the proposed method, we have mined 48 features for each word (24 Mean-squared energies & 24-Mean amplitudes). The effect of Gabor wavelets applied on a word image shown in Fig. 6 is visualized in Fig-7

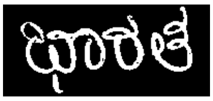
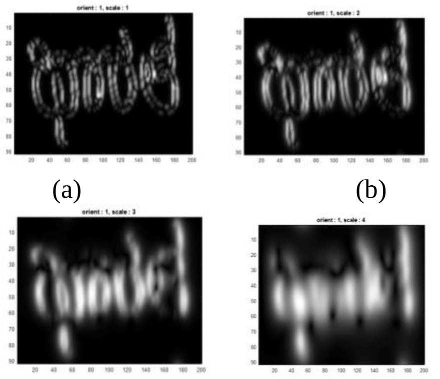


Fig. 6: Sample word image on which Gabor wavelets are applied to visualize its effect shown in Fig.-4.



(a) (b)

(c) (d)

Fig. 7: Visualization of a word image (Figure-3) after applying Gabor wavelets at (a) Scale=1, Orientation=($\frac{n\pi}{k}$) (b) Scale=1, Orientation=($\frac{n\pi}{k}$) (c) Scale=1, Orientation=($\frac{n\pi}{k}$) (d) Scale=1, Orientation=($\frac{n\pi}{k}$)

4.2.2 Histogram oriented gradient (HOG)

HOG stands for Histogram of Oriented Gradient, which is first projected by N. Dalal and B. Triggs [24] in 2005 for human recognition. The simple idea of HOG is that local object appearance and shape can frequently be categorized well by the spreading of local intensity gradients or edge directions, even without exact information of the matching gradient or edge locations. It is employed by dividing the image window into small spatial areas (called cells), for each cell gathering a local 1-D histogram of gradient instructions or edge locations of the pixels of the cell. After the success of HOG in human detection and sketch-based image retrieval [25], J. Almazán et al. [26] have presented HOG into handwritten word spotting problem. In their segmentation-free method, the document images are divided into equal-sized cells and represented by 31-dimension HOG histograms. Queries are characterized analogously using cells of the similar size in pixels. The score of a document region can be projected as the convolution of the query with respect to that area, using the dot product as a similarity measure between the HOG descriptors of each cell. Also, there are some earlier studies uses the histogram of gradients for handwritten word spotting in a similar manner to HOG [27]. For the proposed method we extracted a total of 81 feature set using HOG for better retrieval rate. The below Fig.8 illustrates the HOG feature on a sample handwritten Kannada word image as follows.

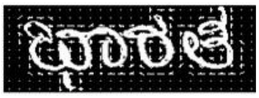


Fig.8 Visualization of HOG features

4.2.3 Local binary pattern

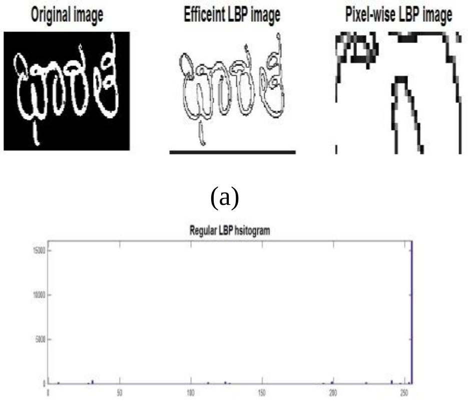
The local binary pattern is an image operator, which converts an image into an array or image of integer values by labeling the small-scale appearance of the image [28].It has confirmed to be highly discriminative and its main points of concern, viz. its invariance to monotonic gray level variations and computational efficiency, make it appropriate for image analysis problem. First, this method was presented by Ojala et al. [29], based on the hypothesis texture that has locally two complementary characteristics, a pattern, and its strength. LBP feature extraction method comprises two main steps: the LBP transformation, and the merging of LBP into histogram illustration of an image. As explained in [29] greyscale invariance is succeeded because of the variance of the intensity of the neighboring pixel to that of the central pixel. It also summarizes the local geometry at each pixel by encrypting binarized variances with pixels of its local neighborhood. In our LBP the unique form of the local binary pattern operator designed by 3x3 pixel block of an image. The pixels of the blocks are threshold by its center pixel value, multiplied by powers of two and then summed to obtain a label for the center pixel. As the neighborhood consists of 8 pixels, a total of 28 = 256 different labels can be obtained depending onthe relative gray values of the center and the pixels in the neighborhood. In the proposed method, we use the uniform LBP as stated by Ojala et.al. [29] Which is the fundamental property of LBP, for the growth of a universal gray-scale invariant operator, the word ’uniform’ in the instance of local binary pattern denotes the consistency of the look i.e. the circular appearance of the pattern has a number of changes. The patterns which are measured as uniform gives a huge majority over 90%, of the 3X3 texture patterns in the historical documents. The most often observed ‘uniform’ patterns matching to important micro-features such as corners, spot, and edges. These are also considered as feature detectors for activating the best similar pattern. In the proposed method where P = 8, and LBP8, Rwill have 256 dissimilar values. And the mathematical representation of the LBP operator as follows.

LBP= $LBP_R^P(Ic) = \sum_{n=0}^{P-1} s(In - Ic)\,2^n$ (5)

Where (In and Ic) are equivalent to the values of center pixel and neighborhood pixel and

$$s(x) = \begin{cases} 1, & x < 0 \\ 0, & x \geq 0 \end{cases}$$

The following Fig. 9(a) and Fig. 9(b) illustrate the effect of LBP and its histogram on handwritten Kannada word is given below.



(a)

(b)

Fig. 9(a) and Fig. 9(b) illustrate the effect of LBP and its histogram on handwritten Kannada word

4.2.4 Morphological features

Image preprocessing based on morphological filters is a group of non-linear processes which is related to the shape or morphology of an image. Therefore, morphological operations trust only on the relative sequence of pixel values, rather than their numerical values, and therefore suited for binary images. In this method we have extracted 20 morphological features for a single word image such as each word image is eroded in four directions viz. 0, 90,180 and -45 degrees. On a word, the opening, top and bottom hat transformations are also employed in vertical and horizontal directions and extracted total 9 features. In addition, the background to foreground ratio of a word is computed after hole filling. To perform these operations, basically we need two images, one is the input image and another is the structuring element. In our case, we made the structuring element based on the average height of characters of a word [30]. Then pre-processed input image I(x,y) is directly used for erosion, opening, top and bottom hat transformations and an output image I0(x; y) is obtained. The arithmetic means of I0(x; y) is defined as

$$Density(\eta) = \frac{\sum_{i=1}^{R}\sum_{j=1}^{C} I'(x,y)}{R * C} \quad (6)$$

where R and C stand for a number of rows and Colom of an image. Beside, hole fill operation is performed on I(x,y) and obtained hole-filled image I0(x; y). Then, its density is computed using equation 1. Thus, 11 features are obtained (erode image density-04, opening, top and bottom hat each-02and foreground, and background information after hole filling) and to make the features independent of font size; they were normalized by dividing them with the size of an image. Further, we added these (09+11) features to become 1x20 feature vector size for the single word image.

4.2.5 Textural filters

Texture filters are also called statistical filters and very robust for characterizing the image by giving information about the local variation of the strength of pixels in an image [31]. In the proposed method the following are the textural filters are extracted on each word image as follows.

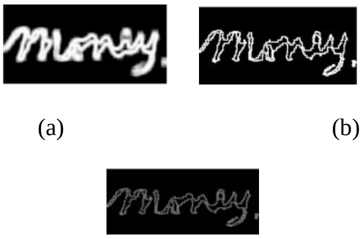
Entropy: This word entropy originates from thermodynamics. It is a statistical quantity of randomness, which is used to describe the texture of the input image. Its value will be high when all the features of the co-occurrence matrix are similar. The cost function of this entropy is given below.

$$Entropy = \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} I(m,n) * \log(I(m,n))) \quad (7)$$

Where m and n are the coefficients of co-occurrence matrix, I(m, n) , is the element in the co-occurrence matrix at the coordinates m and n and N is the dimension of the co-occurrence matrix. The size of this filter is one

Entropy filters: Computes the local entropy of a grayscale image Range filters: Computes the local range of the image.

Standard deviation filters: Calculates the local standard deviation of an image The above three filters yields the output image and its size is similar to the input image; therefore, we computed the mean and standard deviation of the output image. The following Fig 10(a), 10(b) and 10(c) illustrates the effect of the entropy filter, range filters and standard deviation filter on a word image respectively as follows.



(a) (b)

(c)

Fig 10 (a) Effect of entropy filter (b) Effect of range filter (c) Effect of standard deviation filter

1.3 Word image matching

The cosine distance similarity measure has been used in number of handwritten word spotting algorithms and realized its outstanding performance. We employed the same similarity

measure for word spotting in complex datasets. Let us assume two-word images S and R of each length is m. They are characterized by vector

S = ($x_1,\ldots,x_t$) and R = ($y_1,\ldots,y_t$). To define the Cosine distance between these two vectors, a matrix C of dimension m × m is constructed where each element $C_{i,j}$ is the cost function of these two S and R vectors. And the cost function of the distance is given below.

$$Similarty(x,y) = cos\theta = \frac{x.y}{\|x\|.\|y\|} \quad (7)$$

Where x and y are the feature vectors.

1.4 Experimental Environment

The experiments are conducted in two different scenarios, such as learning free and learning based word retrieval. The following Sections provide the detail about both the type of approaches.

1.4.1 Learning free approach:

This method uses similarity measures. There is training and testing datasets. Manually a ground truth is prepared for all the four types of dataset which is given in Section 4.1 and selected 5 different keywords from the dataset. The principle of selection of the query word is that the selected query word should exist at least twice in each document page. The performance evaluation measures used here are mean average precision (mAP), precision, recall, and f-measure. Given a query, we label the set of relevant objects with respect to the query as rel and the set of retrieved words from the database as ret. The precision and recall are defined in terms of ret and rel in Eq. 8 and Eq. 9. precision is the precision at rank rel.

$$Precision(P) = \frac{|ret \cap ret|}{|ret|} \quad (8)$$

$$Recall(R) = \frac{|ret \cap ret|}{|rel|} \quad (9)$$

$$F - Measure(FM) = 2 * \frac{(P*R)}{(P+R)} \quad (10)$$

1.4.2 Learning approach:

Learning-based method is the supervised machine learning method [32] [33] [34], it trains the prototypes of the query words. On the other hand retrieval based method is dedicated to matching process with the query and other words of the corpus set without any related training process [35]. Learning-based approaches are suitable for the problems where the keywords to spot are early known and constant. If the training data is huge and able to deal with several writers, then learning method is preferable. Learning-based approaches [36] [37] uses a statistical approaches to train a keyword prototype that is used to score query words. A common method was recently presented in [36]. Classification methods usually include two steps such as training and testing. In the training step, image properties are extracted and trained to the system to recognize. Further, in succeeding testing, these feature-space partitions are used to classify an image. A vital task in any recognition and computer vision problem is determining the distance between feature vectors of images. Several efforts have been made to describe image distances that give perfect results. Among others, Euclidean distance is the most frequently used distance measures due to its simplicity. Mathematically Euclidean distance is the “ordinary" distance between two points and is given by Eq.11.

$$I(p,q) = I(q,p) = \sqrt{(p_1 - q_1^2) + (q_2 - p_2)^2} \quad (11)$$

Euclidean distance is used in tree classification, medium tree classification and Gaussian support vector machine (GSVM). For tree classifier, the value of k is dependent on the data. Here the value of k is considered as 5 (experimentally fixed). Support Vector Machines are supervised learning methods which are related to learning procedures that observes the data and recognize its patterns .In all three classification process used 5-fold cross-validation (CV) to estimate the performance of the technique. Here, we divided the data into 5 parts. From 5 parts, a single part is used for validation, while remaining 4 parts will be considered for training. This is repeated 5 times for all parts of the data and then all parts of the results are calculated and averaged to get the single value. The cost function the accuracy is given below.

$$Accuracy = \frac{No.of\ Correctly\ classified\ word}{No.of\ Total\ words} \times 100$$

(12)

4.5 Results and discussion

4.5.1 Results on Dataset A, B, C, and D based on similarity test:

In this Section, we discuss the results of word retrieving from the document images. The efficiency of the proposed methods is estimated on four datasets such as Dataset A, B, C and D these are explained in Section 4.1 as George Washington, Kannada handwritten, camera captured and heterogeneous documents respectively. The Dataset A is publicly available [18]. However, Dataset B, C and D are created by us. And the details of these datasets are given in Section 4. The performance of the proposed methods is evaluated through precision, recall, and f-measure as described in section 4.4. Table-2 Illustrates the overall performance of the proposed algorithms with cosine distance measure. These exhaustive experimental results obtained based on various methods exhibiting their potentiality of retrieving the required word from complex datasets. After a keen observation of all the values of performance measuring parameters such as PR and FM of texture filters are high. Features of texture filter are sustaining their performance with all the dataset in terms of PR and FM. Precision recall values with texture features are 81.23, 78.43, 41.99, 60.55 with dataset A, B, C and D respectively.

These are high compared to other methods. Similarly, F-measure values of dataset A, B, C and D are 68.20%, 59.67% , 34.61%, an d 49.11% respectively and these are outstandingly more as compared to other results. The Figure 11 through Figure 17 illustrate the precision and ROC curve for the proposed features sets on a dataset A

Table-2 Results on GW (Dataset A), Kannada Handwritten (Dataset B) and Dataset C based on learning free method (Traditional Word Spotting)

Method	Measures (In average)	Dataset A	Dataset B	Dataset C	Dataset D
Gabor	PR	36.31	33.65	29.44	40.81
	RC	61.65	28.15	21.44	32.71
	FM	44.43	30.36	24.97	35.28
HOG	PR	46.57	42.24	29.64	49.82
	RC	59.10	28.24	20.90	32.35
	FM	51.70	31.36	25.33	39.05
LBP	PR	53.20	45.29	32.26	51.94
	RC	67.72	38.94	22.93	34.61
	FM	58.45	36.28	26.86	41.42
Morphology	PR	34.24	32.23	25.44	45.55
	RC	59.34	30.87	21.44	34.65
	FM	42.45	30.65	23.97	39.55
Comb.1 & 4	PR	39.54	46.29	34.63	51.94
	RC	67.64	30.25	28.95	34.61
	FM	48.77	43.46	27.82	41.42
Comb 2 & 4	PR	53.48	56.92	36.60	56.97
	RC	72.08	37.38	24.97	35.38
	FM	60.36	43.46	30.28	43.52
Comb 3 & 4	PR	61.32	59.61	36.64	50.24
	RC	76.50	37.38	23.91	32.97
	FM	67.13	46.25	29.48	39.51
Texture filters	PR	81.23	78.43	41.99	60.55
	RC	58.78	54.30	28.26	42.10
	FM	68.20	59.67	34.61	49.11

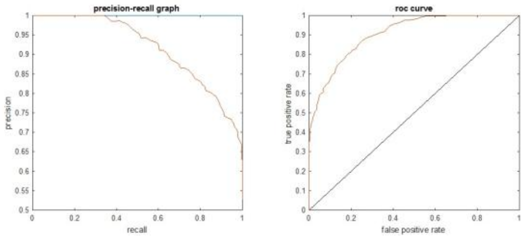


Fig.11 Precision-recall and ROC curve of Gabor features for Dataset A

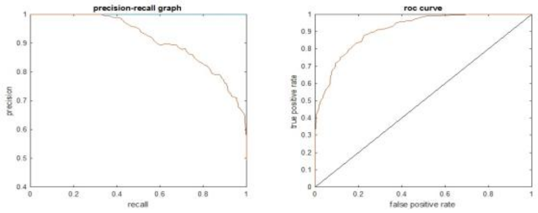


Fig.12 Precision-recall and ROC curve of HOG features for Dataset A

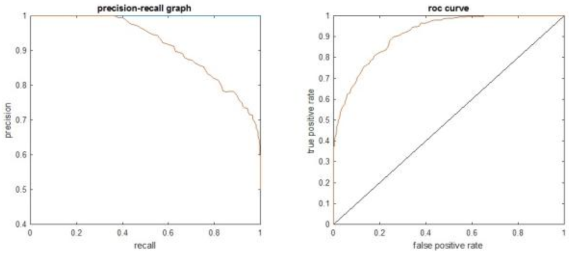


Fig.13 Precision-recall and ROC curve of LBP features for Dataset C

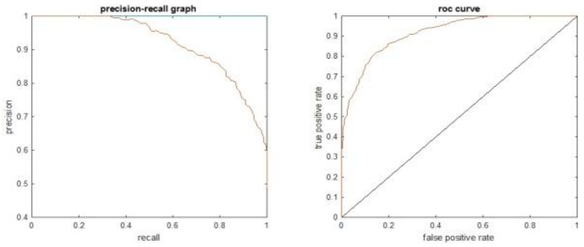


Fig.14 precision-recall and ROC curve of Gabor and morphology features for Dataset A

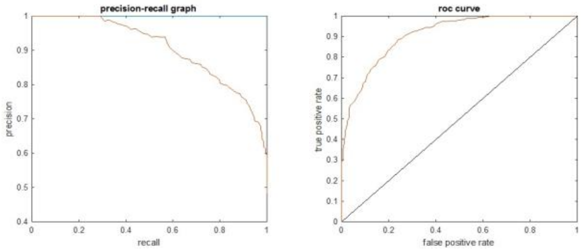


Fig.15 Precision-recall and ROC curve of HOG and morphology features for Dataset A

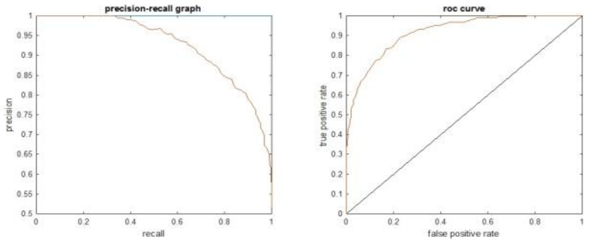


Fig.16 Precision-recall and ROC curve of LBP and morphology features for dataset A

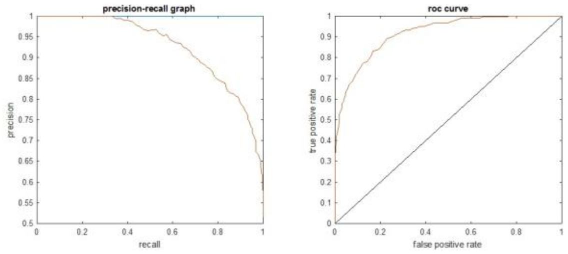


Fig.17 Precision-recall and ROC curve of texture filters for dataset A

4.5.2 Results on Dataset A, B, C, and D based on the learning method:

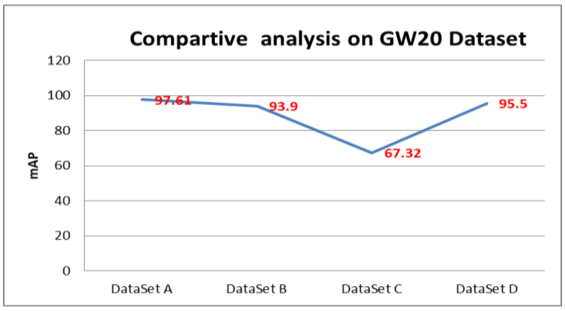
Dataset A: The GW dataset is a public dataset which consists of 20 handwritten pages written by more than one writer belongs to Washington's secretaries. Binarization, segmentation (as discussed above) and normalization of segmented words is performed and obtained 4860 words for experimentation. For training 15 document pages were used. For testing the remaining 5 pages were used. Table 3 shows the results of three classifiers namely tree classifier, medium tree classifier and Gaussian support vector machine (GSVM) on GW dataset with 5-fold cross validation. The mean average precision results are 83.09%, 86.03% and 97.61% with tree classifier, medium tree classifier and Gaussian support vector machine (GSVM) respectively. It is interesting to note that GSVM is outperforming compared to other two classifiers by maintaining balanced accuracy between the classes. In addition, 97.61% mAp is high result compared to all the recent reported works (see Table 4).

On dataset B all the alogrithms porposed above are applied. Here, the experimental setup is formed by making two class problems as query word classes and non-query word class. The words which are occuring more than three times in all the pages are selected as query words. A total of 5 different queries are chosen from 150 pages. In such a way formed five query word classes and each class consist of 80 words. The reaming words are considered as non-query word class. Table 3 presents the results of dataset B with different classifiers. Applied three simple tree classifier, medium tree classifier and Gaussian support vector machine (GSVM) and obtained results as 67.40%, 69.70%, and 93.90% respectively. Mean Average precision (mAp) is used to measure the performance of the proposed methods. It can be noticed that mAp of GSVM classifier is the highest result with accuracy of 93.90%. This experiment endorses outstanding performance even in the case of multi-writer with their different writing styles and age. And further, the experimentation is extended to camera captured document images called Dataset C. The description of this dataset is given the section 4.1. We have selected manually five query words from 50 pages with a condition that a query word should be occurred minimum three times in each page. In this way, five classes of query words forms and each class contains 100 words. The remaining words and considered as non-query words. All the methods which are applied on dataset A and B are extended to C. In this case also the combination of texture features and GSVM classifier are showing high results in terms of mAp (see Table 3) compared to other proposed methods. The mAP of this dataset is 67.32%. It is clear that, there is a sudden fall of 30% results compared to the mAp of Dataset A and B. These results showed the impact of the of the image capture devices and various light variation conditions while capturing the images. The texture classifier with GSVM sustained its performance with such complex datasets.

Dataset D is the mixture of all the above documents. This dataset has been created to know the performance of the proposed algorithms on this heterogeneous dataset and noticed interesting results. Here also texture features with GSVM is giving excellent results i.e. 95.50%. Testing the performance of the algorithms on heterogeneous dataset is essential to confirm that these algorithms work on any group of documents which contains hybrid documents.

Table-3 Results on Dataset A, B, C and D based on the learning method

Method	Classifiers	Dataset A	Dataset B	Dataset C	Dataset D
Gabor	Complex Tree	69.70	55.70	39.40	71.70
	Medium Tree	69.80	57.70	42.20	74.10
	Gaussian SVM	72.56	59.89	47.23	83.60
HOG	Complex Tree	69.80	56.70	41.40	73.80
	Medium Tree	69.90	58.70	43.60	78.30
	Gaussian SVM	72.33	60.87	48.65	86.40
LBP	Complex Tree	65.70	58.80	44.90	76.80
	Medium Tree	66.80	59.30	46.80	81.30
	Gaussian SVM	69.89	66.67	51.56	85.40
Morphology	Complex Tree	56.70	47.60	33.40	66.78
	Medium Tree	56.70	49.70	38.60	68.87
	Gaussian SVM	60.88	55.76	49.78	72.10
Comb.1 & 4	Complex Tree	75.70	66.40	45.40	81.30
	Medium Tree	79.60	62.40	46.70	83.80
	Gaussian SVM	81.52	76.51	55.43	91.40
Comb 2 & 4	Complex Tree	78.60	69.80	43.40	84.10
	Medium Tree	81.30	69.60	44.60	87.30
	Gaussian SVM	90,63	84.55	56.65	94.40
Comb 3 & 4	Complex Tree	78.80	71.40	47.90	83.04
	Medium Tree	76.60	69.50	47.80	84.05
	Gaussian SVM	89.50	82.34	61.56	92.08
Texture filters	Complex Tree	83.90	67.40	45.80	93.50
	Medium Tree	86.30	69.70	47.90	95.40
	Gaussian SVM	97.61	93.90	67.32	95.50



V. COMPARATIVE ANALYSIS

The comparative study is carried out with learning and learning free methods and their analysis is reported in Table 4 and 5 respectively. This comparative analysis is made with GW20 dataset as there are number of similar work exists in the literature. However, we have not made any comparative analysis with the results obtained on dataset B, C and D, because there is no work reported on such datasets. Texture filters with GSVM has been outperforming in all the cases and yielded 97.61% mAP with learning approach (see Table 4). This is the highest result compared to the previously published work. In learning free method 81.23 mAP is obtained with the same combination of features and classifiers (see Table 5). We can clearly notice that texture features with GSVM performance is good with all dataset except dataset C. If we apply an efficient preprocessing on dataset C, certainly results will be enhanced.

Table.4 Comparative analysis of the word retrieval results on George Washington dataset with learning method

Algorithm	mAP	Accuracy (P@1)
Fisher CCA[38] -2014	93.11	95.44
CNN-R-PHOC[14]-2017	79.83	87.71
Co-HOG [40]-2017	91.03	-
CNN-[44]-2018	92.75	
Proposed (Texture filters)	**97.61**	

Table.5 Comparative analysis of the word retrieval results on George Washington dataset with learning free method

Algorithm	mAP	Accuracy (P@1)
HOG descriptors[41]-2014	51.88	-
Fisher Vector [42]-2015	63.87	-
HOG pooled Quad Tree[43] 2016	48.22	64.96
LBP with spatial sampling [43] 2016	54.44	72.86
Proposed (Texture filters)	**81.23**	**90.30**

VI. CONCLUSION

In this paper, we studied various word spotting techniques and their performance with different datasets. In fact, all the algorithms proposed have given inconsistent results when they apply on different datasets. Texture features with GSVM is given 97.1% result with learning approach and it is comparatively higher than all the reported works. But the performance of the same algorithm has been gradually decreasing with respect to dataset B, C and D and it is continued in case of learning free approach. Our study indicating that the algorithms available in the literature are not generic in nature. But our algorithm (texture filters with GSVM) is sustaining its performance with almost all the datasets.

REFERENCES

1. .Manmatha,R.,Han,C., Riseman, E. M., and Croft, W. B., Indexing handwriting using word matching. In Digital Libraries: First ACM International conference. on Digital Libraries, pp. 151-159. 36, 1996.
2. Manmatha, R., Han, C., and Riseman, E. M. Word spotting. A new approach to indexing handwriting. In
3. Proceeding of the Conference. on Computer Vision and Pattern Recognition, pp. 631-637,1996.
4. Manmatha, R., and Croft, W. B. Word spotting: Indexing handwritten manuscripts. In Intelligent Multimedia Information Retrieval,. MIT Press, Cambridge, MA, 1997, pp. 43-64, 1997.
5. Seung-Ho Lee, HyunKyu Lee and Jin H. Kim., On-line Cursive Script Recognition. Using an Island-Driven Search Technique' 0-8186-7128. 1995.
6. Patricia Keaton, Hayit Greenspan and Rodney Goodman,., Keyword Spotting for Cursive Document Retrieval, 0-8186-8055-5, IEEE,1997.
7. A. Kołcz, J. Alspector, M. Augusteijn, R. Carlson and G. ViorelPopescu., A Line-Oriented Approach to Word Spotting in Handwritten Documents, Pattern Analysis & Applications:153–168 2000 2000.
8. T. M. Rath, S. Kane, A. Lehman, E. Partridge, and R. Manmatha., Indexing for a Digital Library of George Washington's Manuscripts: A Study of Word Matching Techniques, 2001.
9. Christophe Choisy, Dynamic Handwritten Keyword Spotting based on the NSHP-HMM, Ninth International Conference on Document Analysis and Recognition (ICDAR) 0-7695-2822. 2007.
10. JoseA.Rodrıguez-Serrano, lorent Perronnin, Handwritten word-image retrieval with synthesized typed queries, 10th International Conference on Document Analysis and Recognition, 2009.
11. Muhammad Ismail Shah Ching Y. Suen, Word Spotting in Gray Scale Handwritten Pashto Documents, 12th International Conference on Frontiers in Handwriting Recognition, 978-0-7695-4221-8/ IEEE 2010
12. .Liang Huang a, FeiYing, Qing-Hu Chen a, Cheng-Lin Liu., Keyword spotting in unconstrained handwritten Chinese documents using contextual word model, Image and Vision Computing 31 958–968, 2013.
13. Marcal Rusinol , David Aldavert, Ricardo Toledo, Josep Llad., Efficient Segmentation-free Keyword Spotting in Historical Document Collections 2015.
14. Leonard Rothacker, Gernot A. Fink., Segmentation-free Query-by-String Word Spotting with Bag-of-Features HMMs, 2015.
15. Suman K. Ghosh, R-PHOC: Segmentation-Free Word Spotting using CNN, 2017.
16. Hafiz Adnan Niaz, UsmanAkram, Usman Akbar., Word Spotting Using Clustering on Extracted DCT and DWT Features, Conference Paper • ICEET1.2018.8338629, 2018.
17. N.Otsu., A threshold selection method from Gray-level histograms. Pattern Analysis and Machine Intelligence. Vol. 9(1), pp. 6266 ,1979.
18. Mallikarjun Hangarge, B.V.Dhandra., Script Identification in Indian Document Images based on Directional Morphological Filters, International Journal of Recent Trends in Engineering, Vol 2, 2009.
19. 19.http://www.fki.inf.unibe.ch/databases/iam-historical-document-database/washington-database.
20. Konidaris T, Gatos B, Ntzios K, Pratikakis I, Theodoridis S, Perantonis SJ,“ Keyword-Guided Word Spotting in Historical Printed Documents using Synthetic Data and user Feedback”,In IJDAR, pp. 167177, 2007.
21. Meshesha M, Jawahar C.V.,Matching word images for content-based retrieval from printed document images, In IJDAR-08, pp. 2938, 2008.
22. T. S. Lee, “Image representation using 2D Gabor wavelets,” IEEE Trans. Pattern Analysis and Machine Intelligence, 18(10), 1996.
23. L. Shen and L. Bai., A review of Gabor wavelets for face recognition, Pattern analysis. Appl. 9: 273-292, 2006
24. Marcolino A, Ramos V, Ramalho M, Pinto JC,“Line and word Matching in Old Documents”, In SIAPR , pp. 123-135, 2000..
25. N. Dalal and B. Triggs., Histograms of oriented gradients for human detection. In Proceedings of the Int. Conf. on Computer Vision and Pattern Recognition, volume 1, pages 886–893, 2005.
26. R. Hu, M. Barnard, and J. Collomosse., Gradient field descriptor for sketch-based retrieval and localization. In Proceedings of the 17th Int. Conf. on Image Processing, pages 1025–1028, 2010.
27. J. Almazán, A. Gordo, A. Fornés and E. Valveny. , Efficient exemplar word spotting. In Proceedings of the British Machine Vision Conference, pages 67.1–67.11, 2012.
28. Y. Leydier, F. LeBourgeois,and H. Emptoz., Text Search for Medieval Manuscript Images. Journal of Pattern Recognition, vol. 40, no. 12, pages3552–3567, 2007.
29. A. Nicolaou, A. D. Bagdanov, M. Liwicki, and D. Karatzas., Sparse radial sampling LBP for writer identification. arXiv preprintarXiv: 1504.06133, 2015.
30. T. Ojala, M. Pietikainen, and T. Maenpaa., Multi-resolution gray-scale and rotation invariant texture classification with local binary patterns,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971–987, 2002.
31. Mallikarjun Hangarge, B.V.Dhandra., Script Identification in Indian Document Images based on Directional Morphological Filters, International Journal of Recent Trends in Engineering, Vol 2 (2009).
32. B.S. Manjunath, and W.Y. Ma., Texture Features for Browsing and Retrieval of Image Data. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18. (8), pp.837-842, 1996.
33. J.Rodriguez-Serrano, F. Perronnin., A model-based sequence similarity with application to handwritten word spotting. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 34, no. 11, pp. 2108–2120, 2012.
34. A. Fischer, A. Keller, V. Frinken, and H. Bunke., “Lexicon-free handwritten word spotting using character hmms,” Pattern Recognition Letters, vol. 33, no. 7, pp. 934–942, 2012.
35. 35.R. Manmatha and T.M. Rath. Indexing of Handwritten Historical Documents Recent Progress, In IEEE SIADIU , pp. 77-85, 2003.
36. R. Manmatha, C. Han, E.M. Risemanl,. Word Spotting: A New Approach to Indexing Handwritings, In CVPR , pp. 631-637, 1996.
37. .Y. Leydier, F. Lebourgeois, and H. Emptoz. Text Search for Medieval Manuscript Images,In Pattern Recognition, pp. 3552-3567, 2007.
38. Cao H, Govindaraju V. Template-Free Word Spotting in Low-quality Manuscripts, In: ICAPR, pp. 4553, 2007.
39. J. Almazan, A. Gordo, A. Fornes, and E. Valveny. Segmentation-free word spotting with exemplar svms, Pattern Recognition, vol. 47, no. 12, pp. 3967–3978, 2014.
40. Thontadari C, Prabhakar C.J. Segmentation Based Word Spotting Method for Handwritten Documents, international Journals of Advanced Research in Computer Science and Software Engineering, Vol-7,Issue-6, 2017.
41. Thontadari C, Prabhakar C.J. Segmentation Based Word Spotting Method for Handwritten Documents, international Journals of Advanced Research in Computer Science and Software Engineering, Vol-7, Issue-6, 2017.
42. J. Almazan, A. Gordo and A. Forn´es and E. Valveny, Segmentation-free Word Spotting with Exemplar SVMs, Pattern Recognition, 2014.
43. S. K. Ghosh and E. Valveny, “A sliding window framework for word spotting based on word attributes,” in Pattern Recognition and Image analysis. , pp. 652–661, 2015.
44. Sounak , Nicolaou, Anguelos, Lladós, Josep and Pal, Umapada.. Local Binary Pattern for Word Spotting in Handwritten Historical Document. 10029. 10.1007/978-3-319-49055-7_51, 2016.
45. Ghosh, Suman & Valveny, Ernest.. R-PHOC: Segmentation-Free Word Spotting using CNN., 2018.

AUTHORS PROFILE

Dr. Mallikarjun Hangarge : He is serving as Vice Principal and IQAC Coordinator at Karnatak Arts, Science and Commerce College, Bidar. He has completed his masters and Ph D from Gulbarga University in 1989 and 2009 respectively. He has published more than 75 research articles in reputed journals and conference proceedings. He has been guiding 8 PhD students. He has completed three projects of Rs. 15 lakhs with the financial assistance of UGC. He had been to USA and Hong Kong for paper presentation with the travel grant of UGC and IAPR (International Association of Pattern Recognition). He has been awarded SRF in 2012 by Indian academy of Sciences, Bangalore. He has collaboration with University of South Dakota, USA, Computer Vision and Pattern Recognition Unit, Indian Statistical Institute Kolkata and Speech Processing Laboratory, IIIT Hyderabad. Dr. Hangarge's research interests are in Computational Intelligence and Pattern Recognition and its applications such as Automatic Handwriting Analysis, Document Image Processing, Natural Language Understanding and Multimodal Biometrics etc. He has been delivered more than 30 invited talks at various national and International Conferences and workshops. He serves on Editorial board of 6 International Journals and Intentional Conferences from USA, Republic of Macedonia, Malaysia, Czech Republic Ostrava and Singapore etc. He is Member of IAENG, Hong Kong,

Senior Member of IACSIT Singapore, Computer Science Teachers Association, USA and Member of Internet Society, Switzerland. He is an ambassador of NPTEL, IIT Madras and Spoken tutorial of IIT Mumbai to publicize these programmes in Hyderabad Karnataka Region.

Veershetty Chitaguppa:

He received the M.Sc. degree in Computer Science from the Gulbarga University, Kalabuaragi India, in 2005, M. Phil in Computer Science from the Madurai Kamaraj University Madurai Tamilnadu, in 2007 and presently pursuing his Ph.D. degree in Computer Science from Gulbarga University, Kalabuaragi under the guidance of Dr.Mallikarjun Hangarge. Presently he is working as Assistant Professor at GFGC Basavakalayan, India. His current research interests include Document Image Analysis and Pattern Recognition. He has published more than 10 research articles in leading journals, conference proceedings indexed at Scopus database. He holds several professional designations including BOE Chairman, and members. He is also a life member of KGCTA Karnataka.