Expert Systems With Applications 76 (2017) 113–128

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Keyword spotting in doctor's handwriting on medical prescriptions

Partha Pratim Roy[a,*], Ayan Kumar Bhunia[b], Ayan Das[b], Prithviraj Dhar[c], Umapada Pal[d]

[a] Department of CSE, Indian Institute of Technology Roorkee, India
[b] Department of ECE, Institute of Engineering & Management, Kolkata, India
[c] Department of CSE, Institute of Engineering & Management, Kolkata, India
[d] CVPR Unit, Indian Statistical Institute, Kolkata, India

ARTICLE INFO

Article history:
Received 6 February 2016
Revised 2 December 2016
Accepted 24 January 2017
Available online 31 January 2017

Keywords:
Word spotting
Printed/handwritten text separation
Knowledge extraction
Hidden Markov model
Medical prescriptions

ABSTRACT

In this paper, we propose a word spotting based information retrieval approach for medical prescriptions/reports written by doctors. Sometimes due to almost illegible handwriting, it is difficult to understand the medication reports of doctors. This often confuses the patients about the actual medicine/disease names written by doctors and as a consequence they suffer. A medical prescription is generally partitioned into two parts, a printed letterhead part containing the doctor's name, designation, organization name, etc. and a handwritten part where the doctor writes patient's name and report his/her findings and suggests medicine names. There are many significance impacts of the proposed work. For example, such work can be used (i) to develop expert diagnostic systems (ii) to extract information from patient history that can be obtained by this proposed method (iii) to detect wrong medication (iv) to make different statistical analysis of the medicines prescribed by the doctors etc. To extract the information from such document images, first we extract the domain specific knowledge of doctors by identifying department names from the printed text that appears in letterhead part. From the letterhead text, the specialty/expertise of doctors is understood and this helps us to search only relevant prescription documents for word spotting in handwritten portion. Word spotting in letterhead part as well as in handwritten part has been performed using Hidden Markov Model. An efficient MLP (Multilayer Perceptron) based Tandem feature is proposed to improve the performance. From the experiment with 500 prescriptions, we have obtained encouraging results. Information from printed letterhead part improved the word spotting performance in handwritten part, significantly.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Automatic interpretation of handwritten documents has been one of the popular research areas in the last few decades due to its large scope of practical applications. Handwriting recognition is challenging because of the large variability of writing styles, cursive nature, vocabulary size, etc. (Bunke, 2003; Liu, Koga, & Fujisawa, 2002). Though high recognition performance is achieved in isolated numeral/character recognition, offline word recognition is not easy. Due to such challenges, the existing recognition systems cannot meet the requirements of real applications. Though the techniques in this field have been used in applications such as automatic address reading (Niyogi, Srihari, & Govindaraju, 1996; Srihari, & Kuebert, 1997; Srihari, Huang, & Srinivasan, 2005), bank check processing (Jayadevan, Kolhe, Patil, & Pal, 2012), and recognition of text filled on forms by hand (Milewski, Govindaraju, & Bhardwaj, 2009). It is rarely applied to doctors' handwritten prescription because of great challenges in readability using the naked eye (Chen, Gong, Li, Tan, & Pang, 2010). To the best of our knowledge, information retrieval in offline prescription images was not taken care earlier although it has significance impacts. For example, such work can be used (i) to develop expert diagnostic systems (ii) to extract information from patient history that can be obtained by this proposed method (iii) detection of wrong medication (iv) to analyse different medicines prescribed by the doctors etc.

While checking patients, doctors write reports and prescribe the medicines. Nowadays, in many hospitals, printed medical prescription formats are used which is easy to keep records of these datasets as well as for analysis of various medicines prescribed by practitioners across the world. The statistical analysis of medicines provides a good idea about that particular brand of medicine which is most widely preferred as well as those which are of less preference. Difficulty arises when handwritten prescription evaluation comes into play as traditional OCR (Optical Character Recognition) of such documents will fail most of the times.

* Corresponding author.
E-mail addresses: proy.fcs@iitr.ac.in, 2partharoy@gmail.com (P.P. Roy).

http://dx.doi.org/10.1016/j.eswa.2017.01.027
0957-4174/© 2017 Elsevier Ltd. All rights reserved.

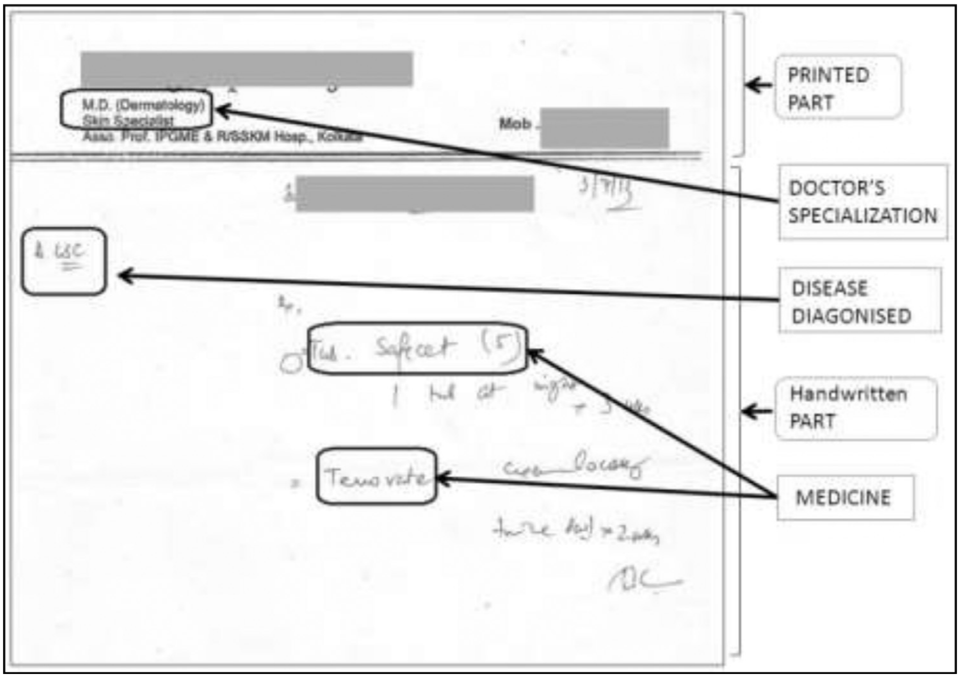

Fig. 1. An example of a prescription from our dataset. Arrows point to different medicine ("safecet", "tenovate") and disease ("LSC") names written by the doctor. Department name ("Dermatology") is mentioned in the printed letterhead part. To keep confidentiality, doctor's name, mobile number, etc. are deleted here.

A medical prescription is usually a letterhead of personal/organizational template, partitioned into two parts: a printed letterhead part at the top portion of the document containing the doctor's name, designation, doctor's registration number, organization name, etc. and a handwritten part where doctors write the patient's name and report medical diagnosis and prescribe medicine names. Fig. 1 shows an example of medical prescription. The medicine names and their intake regime are written to advise the patients. Next, patients read the prescription and take medicines accordingly. Often the patients face serious problems to understand the report because of the illegible handwriting of doctors and unstructured text. A misread prescription can lead to mistreatment which can seriously affect the patients and quality of healthcare.

The objective of this paper is aimed at automatic transcription of handwritten prescription (which poses great difficulties in readability) with the help of keyword spotting. As mentioned earlier, to the best of our knowledge, such work on doctor's prescriptions was not taken care earlier. Thus, the problem itself is novel. Moreover, the proposed work has many significant impacts as mentioned earlier.

The contribution of this paper is two-fold. First, we propose an approach by incorporating domain knowledge into text retrieval process which is another novelty of the work. From the printed letterhead part of the prescription, we retrieve the attributes regarding the doctor specifications (discipline/ department/ designation) with the help of word spotting in printed text part. The doctor's domain thus extracted is indexed with the search of disease/medicine keywords. It utilizes a handwritten word spotting methodology to search the queried keywords in the written text portion of the relevant prescription. Second, a novel Tandem-HMM system is proposed for word spotting. From the experiment it is noted that Tandem approach outperforms traditional way of word spotting schemes.

Fig. 2 explains the overall idea of our word spotting system. The traditional handwriting recognition systems will not be able to recognize the full writing due to complex handwriting challenges of the doctors' writing style. Thus, word-spotting approach

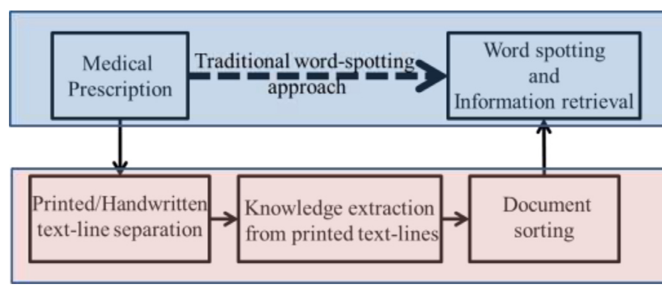

Fig. 2. The word spotting approach considered for information retrieval in prescription images.

is developed to improve the word searching without recognition. Word spotting in letterhead part as well as in handwritten part has been performed using Hidden Markov Model. To improve the performance, an efficient MLP based Tandem feature is integrated. Information from printed letterhead part improved the word spotting performance in handwritten part significantly.

2. Related work

Word spotting techniques have been popular in speech (Rose & Paul, 1990) and handwriting (Bluche, Ney, & Kermorvant, 2013; Rath & Manmatha, 2007) recognition community. The advantage of word spotting is that it allows users to search for keywords on-the-fly in the document without transcribing the handwriting. There exist many pieces of work on word spotting in applications of postal document (Niyogi et al., 1996; Srihari & Kuebert, 1997), bank check (Jayadevan et al., 2012), digital libraries (Nagy & Lopresti, 2006), historical documents (Antonacopoulos & Downton, 2007), etc. Word spotting, which was initially proposed by Rath and Manmatha (2007), usually refers to finding or matching a query keyword from a text line. Dynamic Time wrapping (DTW) has been used (Rath & Manmatha, 2007) with conjunction of profile features. Though DTW can be used to provide similarity score of two word images, it is computationally slow and may not work in

Table 1
A summary of related work on word spotting.

Reference	Method	Segmentation
Rodríguez-Serrano and Perronnin (2012).	Semi-continuous HMMs	Words
Almazán et al., 2014a.	Pyramidal histogram of characters	Words
Frinken et al. (2012).	Recurrent Neural Networks	Lines
Fischer, Keller, Frinken, and Bunke (2012).	Character HMMs	Lines
Wshah et al. (2014)	Statistical script independent word spotting	Lines
Almazán et al., 2014b.	Exemplar SVM with re-ranking	None
Leydier et al. (2009)	Gradient features with elastic distance	None

complex word images. Mainly, two different approaches of word spotting exist, 'Query By Example (QBE)' (Rath & Manmatha, 2007; Roy, Ramel, & Ragot, 2011; Rusinol, Aldavert, Toledo, & Lladós, 2015) that searches using a given word image whereas 'Query By Text (QBT)' (Rodríguez & Perronnin, 2009) uses a query string.

The QBE approach (Rath & Manmatha, 2007) is usually performed at the full image level using a set of features at different zones of the word image. Different features are explored in the literature based on the dataset of experiment. "Gradient, Structural and Convexity" (GSC) features (Zhang, Srihari, & Huang, 2003), Gabor features (Cao & Govindaraju, 2007), Corner features (Rothfeder, Feng, & Rath, 2003), "Gradient Angular Feature" (GAF) (Leydier, Ouji, LeBourgeois, & Emptoz, 2009) are some of them. In historical manuscript, zones of interest was explored in order to match informative parts of the words (Leydier et al., 2009). A more discriminative feature of word images was presented by exemplar SVM framework in Almazán, Gordo, Fornés, and Valveny (2014b). However, problems are reported for keywords not occurring in the training set. These techniques restrict users to search a list of fixed keywords that were trained before.

However, QBT has been proved to be better than QBE approach (Rodríguez & Perronnin, 2009). These approaches are trained to search query keyword both in word and line levels. Different features have been explored by researchers. Recently, Hidden Markov Model based approach (Wshah, Kumar, & Govindaraju, 2014) has been applied in text lines for finding the query text word. The features in HMMs follow a sliding window approach: a fixed width window shifts from left-to-right. At each location of the window, a feature vector is extracted (Rodríguez & Perronnin, 2009) and the sequence of feature vectors obtained in this fashion is modelled with Semi-continuous HMM. This approach has been tested only in segmented words. A pyramidal histogram of character based training approach was proposed (Almazán, Gordo, Fornés, & Valveny, 2014a) for improved word retrieval performance. The discriminative Recurrent neural network is also explored to search the keyword in text line (Frinken, Fischer, Manmatha, & Bunke, 2012). In our system, we have improved the traditional HMM based word spotting with discriminative tandem features which make it more efficient in line level word spotting. A summary of some of these QBE and QBT based approaches are tabulated in Table 1.

The rest of the paper is organized as follows. We describe the keyword spotting approach used in our system in Section 3. In Section 4, we explain the knowledge extraction from letterhead part of prescription and integrate the process in word spotting in handwritten text. Section 5 demonstrates data details and experiment details of our system. Finally, conclusions and future work are presented in Section 6.

3. Keyword spotting

We consider word spotting for information retrieval from printed as well as handwritten part of prescription images. The keywords will be detected from prescription line images using Hidden Markov Models (HMMs).For matching purpose, a score, that represents the likelihood of the query keyword, will be assigned to each target image. Based on the likelihood score, the image is returned as a positive match. Details of feature extraction approach and word spotting are discussed in the following subsections.

3.1. Hidden Markov model

In the field of handwritten text recognition, Hidden Markov Models have been extensively used because of its efficiency of recognition in the cases of touching characters, distorted characters even without proper pre-processing (Marti & Bunke, 2002). The basic model is the character HMM which consists of J hidden states (S_1, S_2 … S_J) in a linear topology as an observation O where ith observation (O_i) represents an n-dimensional feature vector x modelled using a Gaussian Mixture Model (GMM) with probability $P_{S_j}(x)$, $1 < j < J$ given by

$$P_{S_j}(x) = \sum_{k=1}^{G} W_{jk} N(x|\mu_{jk,} \Sigma_{jk})$$

Where G is the number of Gaussians and N refers to a multivariate Gaussian distribution with mean μ_{jk}, covariance matrix Σ_{jk} and probability W_{jk} for kth Gaussian in state j.

For training the model, firstly, feature vector is extracted from labelled text line images. The probability of the character model of the text line is then maximized by Baum-Welch algorithm assuming an initial output and transitional probabilities. Using the character HMM models, a filler model has been created which is shown in Fig. 3(a). Fig. 3(b) shows the keyword model which has been used in our system to spot a keyword in a text line image. The filler model represents a single character model consisting of any one of the characters among 'Char i's, where $1 \le i \le N$ (see Fig. 3(a)). A 'Space' model has been used in the keyword model shown in Fig. 3(b) which is accounted for modelling white spaces.

3.2. Feature extraction

Pyramid Histogram of Oriented Gradient (PHOG) (Zhang et al., 2003) is the spatial shape descriptor which gives the feature of the image by spatial layout and local shape, comprising of gradient orientation at each pyramid resolution level. PHOG feature has been found efficient in word recognition system (Bhunia, Das, Roy, & Pal, 2015). To extract the feature in each sliding window, we have divided it into cells at several pyramid level. The grid has 4^N individual cell at N resolution level (i.e. N=0, 1,2,..). Histogram of gradient orientation of each pixel is calculated from these individual cells and is quantized into L bins. Each bin indicates a particular octant in the angular radian space. The concatenation of all feature vectors at each pyramid resolution level gives the final PHOG descriptor. L-vector at level zero represents the L-bins of the histogram at that level. At any individual level, it has $L \times 4^N$ dimensional feature vector where N is the pyramid resolution level (i.e. N=0, 1, 2….). So, the final PHOG descriptor consists of $L \times \sum_{N=0}^{N=K} 4^N$ dimensional feature vector, where K is the limiting

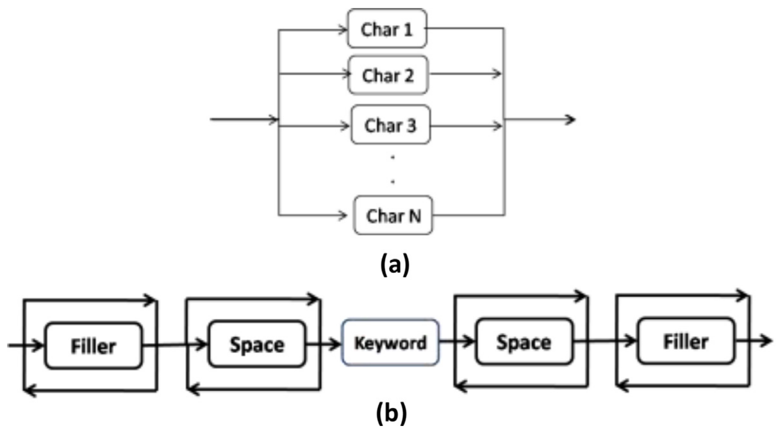

Fig. 3. (a) Proposed Filler Model and (b) Keyword Model.

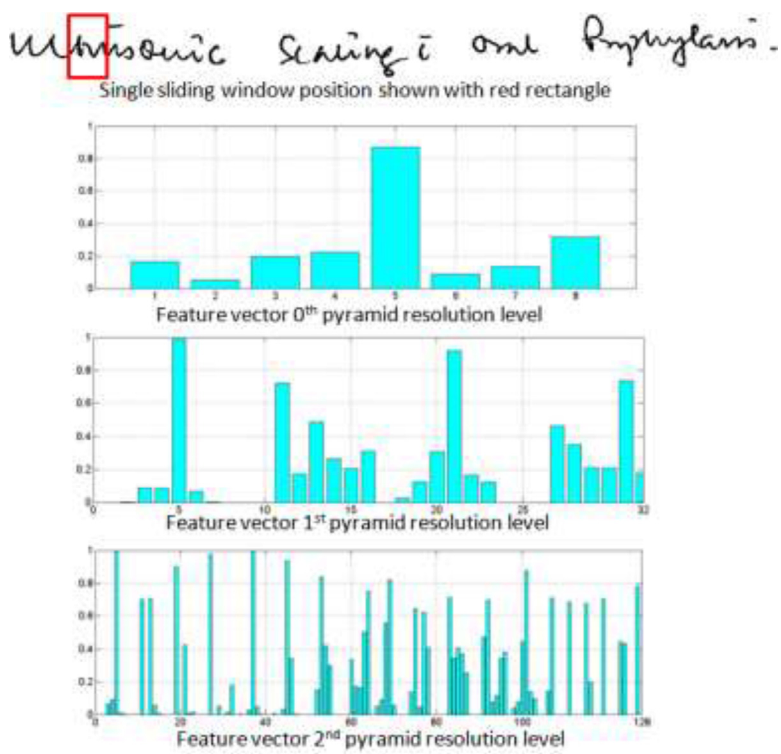

Fig. 4. Extraction of PHOG features in a sliding window patch from text line.

pyramid level.Here we have limited the level to K = 2 in our implementation and we considered 8 bins 360°/45° of angular information. So we got $(1\times8)+(4\times8)+(16\times8)=(8+32+128)=168$ dimensional feature vector for individual sliding window position. An example of feature extraction in a text line is depicted in Fig. 4.

3.3. Tandem feature

We present an efficient Tandem-PHOG feature for word spotting in our system. Tandem modelling (Hermansky, Ellis, & Sharma, 2000) was proposed to combine the discriminative power of the Artificial Neural Networks with the sequence modelling ability of the HMMs in speech recognition. The positive effect of the combined feature is that the MLPs performa non-linear feature transformation into a space that is explicitly oriented for discriminability of characters/states. The transformed feature leads to improved discrimination by the GMM which describes the output space associated with each HMM state. The advantage of the tandem approach is due to its robustness to noise.

We present the MLP-HMM tandem systems for our word spotting purpose in Fig. 5. For this purpose, MLPs are integrated into the HMM framework to form the tandem MLP-HMM system. In contrast to the likelihood of a feature vector $x_{t,}$ given an arbitrary state s_i, MLPs produce state posterior probability $P(s_i \mid x_t)$. Training the MLP requires each observation at time step t in the training data to be aligned to a character label in the word. However, the class (e.g., HMM-state) labels are usually not available. To do so, a previously trained GMM-HMM is applied to the training data in the forced alignment mode. Next, the MLP is trained on the labelled observations in a frame-based approach. The trained MLP is used to compute the posterior distribution over the character labels for each observation. To perform the sequence modelling in a tandem HMM approach, the posterior estimates are considered as observations to train a new HMM (GMM-HMM). Generation of Tandem feature is illustrated in Fig. 5. The output posterior

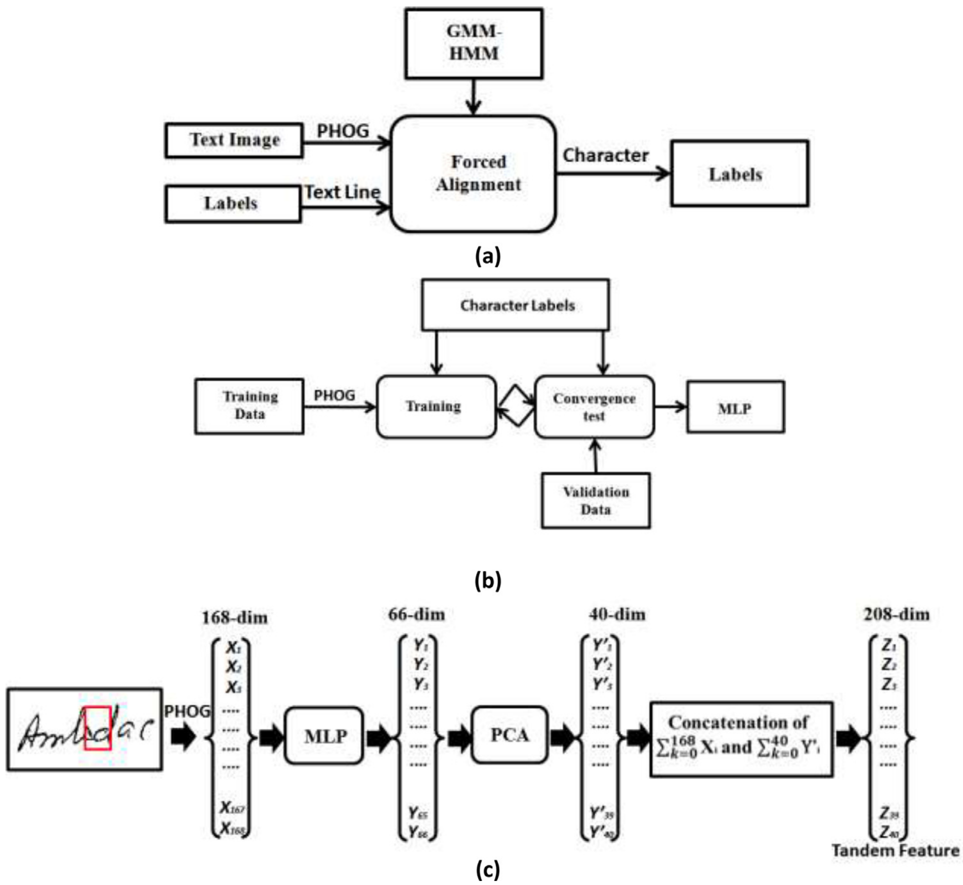

Fig. 5. Tandem feature preparation from text line image. (a) Character labels are obtained using Viterbi Forced Alignment. (b) Sliding window features of character are trained using MLP classifier. (c) Sliding window feature and MLP feature are concatenated to feed into HMM.

probabilities are decorrelated by a dimensionality reduction algorithm. We applied Principal Component Analysis (PCA) (Bishop, 2009) to the posterior probabilities of the MLPs. It is done to reduce the dimensionality and to orthogonalize the feature vectors. In MLP, the learning rate of 0.01 was considered in our system. Next, the features were normalized by mean and variance. Finally, these reduced feature vectors were concatenated with the baseline HMM features.

3.4. Text line scoring

Word spotting mechanism is based on the scoring of a text image for the keyword. The left-to-right Bakis topology is used in HMM structure. Given a sequence of sliding window features (tandem feature) from text-line image, the HMM is trained to compute the likelihood of a query keyword in that line. Next, a score is computed from the likelihood score. If the score value is greater than a certain threshold, then it gives a positive value for the occurrence of that particular keyword in that text line. The score assigned to the text line image X for the keyword W is based on the posterior probability $P(W|X_{a,b})$ trained on keyword models. Here a and b correspond to starting and ending position of the keyword whereas $X_{a,b}$ gives the particular part of text line containing the keyword (Frinken et al., 2012). Applying Bayes' rule we get

$$\log p(W|X_{a,b}) = \log p(X_{a,b}|W) + \log p(W) - \log p\left(X_{a,b}\right)$$

Considering equal probability we can ignore the term $\log p(W)$. The term $\log p(X_{a,\,b}|W)$ represents the keyword text line model where it is assumed that exact character sequence of the keyword to be present separated by 'Space'. The rest part of the text line is modelled with Filler text line model. Then we can find the position a, b for the keyword alongside with the log-likelihood; $\log p(X_{a,b}|W) = \log p(X_{a,b}|K)$.

$\log p(X_{a,\,b})$is the unconstrained filler model F.The general conformance of the text image to the trained character models is obtained by log-likelihood$\log p(X_{a,b}) = \log p(X_{a,b}|F)$. The difference between the log-likelihood value of keyword model and filler model is normalized with respect to the length of the word to get the final text line score.

$$Score(S, W) = \frac{(\log\ p(Xa,b\ |\ K) - \log\ p(Xa,b\ |\ F))}{b - a}$$

Then this *Score*(*S, W*) is compared with a certain threshold for word spotting.

4. Knowledge extraction and information retrieval

Before applying the word spotting in text lines, we have extracted the text line images from the prescription as follows. The document image is first converted into binary image using global histogram-based binarization method. Our word spotting approach searches the keyword in binary text word images. Hence, the binary document is segmented into individual text lines using a line segmentation algorithm (Roy, Pal, & Lladós, 2008). Here, some seed components of a line are obtained from smoothed text regions of document. The upper and lower boundary information of a text line is obtained from background regions. Next, foreground seed components and boundary information are used to segment the text-lines.

After extracting the text lines, the text mode of the lines is identified as printed or handwritten using HMM based approach.

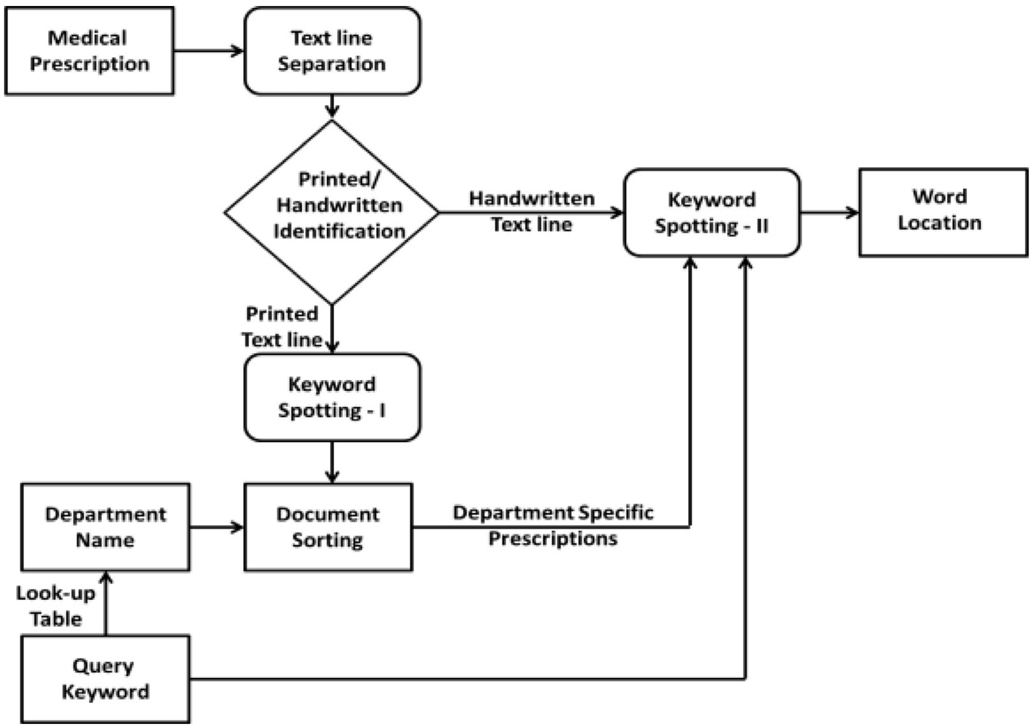

Fig. 6. The flowchart of the proposed word spotting system in medical prescriptions. Keyword Spotting-I deals with printed text lines, and Keyword Spotting-II handles handwritten text lines for information retrieval.

The query keywords such as medicine/disease names which are used to search handwritten portion of prescription images, are considered to find the department names first from the prescription using a Look-up table. The documents are next sorted using word spotting by the department name in printed text (or letterhead) part. Finally, only the handwritten lines obtained from specific department are searched with the query keywords. The flowchart of this knowledge propagation and word spotting is illustrated in Fig. 6. Details of these modules are discussed in the following subsections.

4.1. Identification of printed and handwritten text

There exist many pieces of work on handwritten/machine printed text identification where the classification is performed at text line, word, or character level (Wshah et al., 2014). Usually, machine printed text words are arranged in a regular fashion with a straight baseline, while handwritten text is irregular with varying baseline. Hidden Markov Model based classifier has been applied successfully for line/word level handwritten/machine printed text classification (Cao, Prasad, & Natarajan, 2011; Guo & Ma, 2001). Our system also follows the similar HMM-based printed/handwritten line identification. Sometimes the line segmentation algorithm (Wshah et al., 2014) fails to join all words in a line and some lines comprise of only one word. Since our word spotting approach needs to handle such single word data, we need to identify the class (printed/handwritten) of the word. Hence, in our system we apply HMM based approach at line and word level classification. The left-to-right Bakis topology has been used in HMM modelling. The sliding window based feature extraction is applied after normalizing the text height. For this purpose, a sliding window is traversed through the line/word to extract the writing-type information. Next, PHOG (Pyramid Histogram of Oriented Gradients) features, described earlier, are extracted in each sliding window. The PHOG feature is used as observation of writing type from the internal state of HMM. Finally, the feature sequence obtained from sliding window is fed to HMM classification. The HMM predicts the modality of the text word whether it is machine printed or handwritten (see Fig. 7).

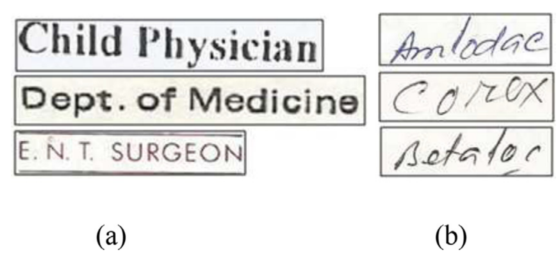

Fig. 7. Some examples of (a) machine-printed and (b) handwritten text words separated by our system.

4.2. Knowledge propagation and document sorting

As mentioned earlier, while dealing with the handwritten part to retrieve the disease/medicines, the information from printed letterhead part is used to narrow down the target documents. The letter-head part of prescription is easier to search because generally it contains information about the doctor's name, organization, specialization in printed text. The query keyword (disease/medicines) is thus mapped to keywords from letter-head part.

A Look-Up-Table (LUT) was created for linking the keywords of medicine/disease names with their corresponding department, i.e. the department of medical science under which the disease is categorized in. The LUT stores the information of particular department of a corresponding query keyword. For example, if we search a medicine e.g., 'Amlodac', which is generally prescribed by the doctors when an individual has 'Heart Failure', it lies under the department of 'Cardiology' (see Table 2). Thus while searching the medicine name taken as user keyword, the process is developed to sort out with respect to its reference disease along with the

Table 2
Some keywords from our prescription dataset. The table shows the disease and medicine names corresponding to the departments mentioned in prescriptions.

Department	Disease	Medicine	Department	Disease	Medicine
Cardiology	Heart Failure	Amlodac Cartirex	E.N.T.	Allergy	Ambrodil Claric
Dental	Periodontitis	Hifenac Lancet	Orthopaedics	Fracture	Calphos Estovon
Dermatology	Cellulites	Aziderm Microdox	Paediatric	URTI	CepodemXP Wysolone

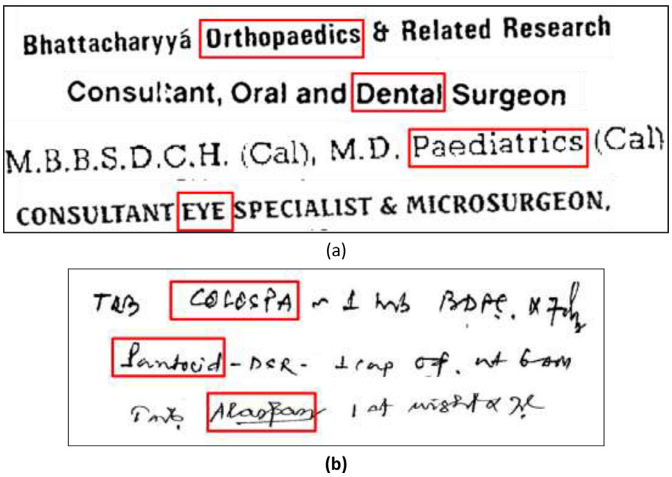

Fig. 8. Examples of word spotting in (a) printed text lines using query words 'Orthopaedics', 'Dental', 'Paediatrics', 'EYE', respectively, (b) handwritten text lines using query words 'COLOSPA, 'Pentocid', 'Alaspan', respectively in handwritten text.

corresponding department and hence the 'department name' is the search key in the resultant files where the 'department name' is actually the doctor's specification that has already been recognized.

4.3. Word spotting in relevant documents

For a particular query keyword, we search it in our look-up table and get the department specification of the query keyword. Thus, we check only our query word for a particular department rather than checking all prescriptions. In Fig. 8(a), some results of word spotting using department name (query word) are shown. After extracting the relevant document using printed text information we perform word-spotting for query keywords only for prescriptions corresponding to particular department. Thus, the system searches the medicine/disease names and refers to the corresponding prescriptions available in the dataset, in which the keywords can be found and now word spotting can be done on the referred files. An algorithm describing the whole system is described in Algorithm 1. Fig. 8(b) shows word spotting results in handwritten lines.

5. Experiment results and discussion

5.1. Document collection & performance protocol

To the best of our knowledge, there exists no standard database to evaluate information extraction in doctors' handwriting. Medical prescriptions are not easily available. Moreover doctors write the prescriptions and hand it over to the patient and rarely do they keep a copy of it. Prescription contains the medical details of an individual which in terms of medical professionals is a highly confidential entity. For our experiment, we have collected 500 prescriptions written by doctors. These were collected from 230 patients

Algorithm 1 Word spotting in medical prescriptions.

Require: Query Keywords (Q_k) with Department Specification (Q_d) and Segmented Text Lines from Prescriptions
Ensure: Detection of the query keyword in Prescriptions
Step 1: Separate printed and handwritten lines using script identification (Section 4.1)
Step 2: From printed text lines find Department Specification (Q_d) using keyword spotting approach.
(Section 4.2)
If Q_d is found
Step 3: Search for Q_k in handwritten lines corresponding to that prescription of specific department using keyword spotting approach. (Section 4.3)
else
Step 4: Do not search in handwritten lines of that prescriptions.
end if

of different profession. We have checked that a total of 18 fonts are used in the letterhead. Some of these fonts are cursive and appear similar to handwriting. The documents contain handwritten text lines ranging from 6 to 25.Some prescription images are made publicly available[1] for further research in this area. In this dataset, standard names of medicines have been used by doctors. The documents were scanned at 300 dpi.

A wide variety of prescriptions have been collected which includes different departments like orthopedics, dental, medicine, ENT, cardiology, eye, pediatrics, dermatology and gynecology. There appear diseases like peridontitis, fever, fracture, heart attack, chicken pox, infections, diarrhoea, pregnancy, URTI, etc. Some of the medicine names like, metrogyl, hifenac, lancet, calpol, pacimol, crocin, augmentin, etova, estovon, panedo, asfer, asilac, cifran,etc. Following experiments are conducted to evaluate the efficiency and effectiveness of the proposed method. The present study includes 11 department names (doctor's specialization), 25 disease names and 145 medicine names.

We have measured the performance of our word spotting system using precision, recall and mean average precision (MAP). The precision (P) and recall (R) are defined as follows.

$$Precision = \frac{TP}{TP+FN} \quad Recall = \frac{TP}{TP+FP}$$

Where, TP is true positive, FN is false negative and FP is false positive. MAP value is evaluated by the area under the curve of recall and precision.

5.2. Experiment on printed/handwritten text separation

For our experiment, text lines are first segmented from the prescription images as discussed in Section 4. After extracting the text lines, the text lines in printed header (Letterhead) parts and that in handwritten parts are separated. The printed and handwritten text lines are classified into two classes by using HMM. We have used PHOG feature for this purpose. We performed our

[1] https://sites.google.com/site/2partharoy/dataset

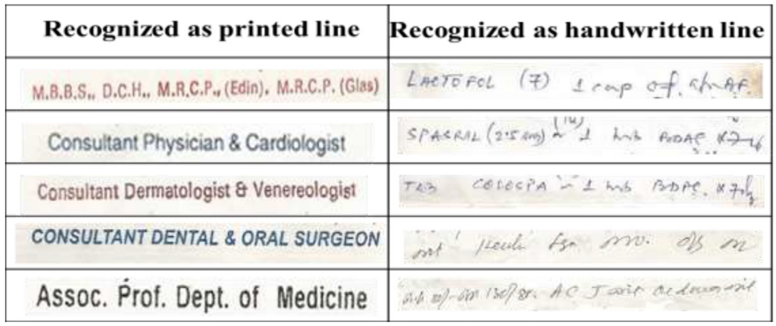

Fig. 9. Some examples of qualitative results where our system identified the text lines correctly.

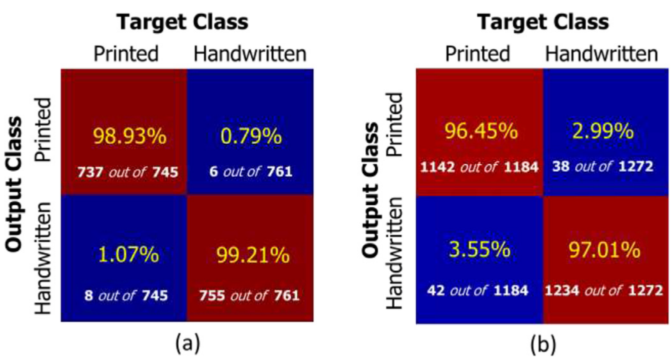

Fig. 10. Confusion matrix for (a) line level classification and (b) word level classification.

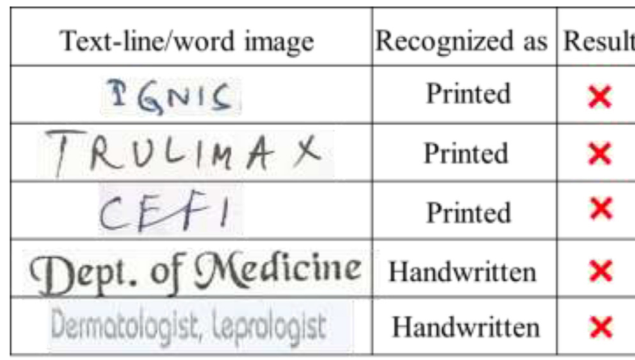

Fig. 11. .Some qualitative results where the method did not work.

printed/handwritten separation experiment on 2152 samples (of both types) as training data and 1506 samples as testing data. We achieved an identification accuracy of 99.1% using PHOG features. Fig. 9 shows some qualitative results where our approach identifies the lines correctly. For word level identification, we have tested using 2456 word level images (printed and handwritten) and achieved an accuracy of 96.7%. The confusion matrices in line and word levels are shown in Fig. 10. With Tandem-PHOG feature, we obtained similar performance.

To compare our printed/handwritten line and word identification approach, we have tested with popular LGH (Local Gradient Histogram) feature (Rodríguez & Perronnin, 2009). Similar to PHOG, a sliding window is being shifted from left to right of the word image with an overlapping between two consecutive frames. Each sliding window patch is next divided into 4×4 cells and from each cell, histogram of gradient (with 8 bins) is computed. The size of the feature dimension is 128. PHOG feature provided an improvement of 0.24% accuracy over LGH in word level identification. The line level accuracy using LGH feature remained same. The identification experiment was also performed with Tandem features but the performance did not improve much.

5.2.1. Error analysis

Fig. 11 shows some examples where HMM-based approach failed to identify correctly. We noted that, the text lines which do not contain much cursive characters are prone to such errors. Similarly, the printed words which appear like cursive text got affected much by this approach.

5.3. Experiment on word spotting in printed text lines

After identifying the text lines/words, we apply word spotting approach to printed text lines to categorize the text lines according to medical department names. For this experiment, we collected a total 5125 words from newspaper image and images from books for training. For testing purpose, we collected a total of 1121 printed text line images from the letter head portion of the prescription. Fig. 12 shows some qualitative results using query words with Tandem PHOG features.

Fig. 13 shows the Precision and Recall curve using PHOG feature and PHOG-Tandem features. In our experiment, 50 keywords have been considered to measure the performance analysis and comparative studies. For comparison, we have considered LGH (Rodríguez & Perronnin, 2009) feature for word spotting. Precision-Recall curve with LGH, PHOG and Tandem-PHOG features are shown in Fig. 13(b).

We have evaluated the performance for word spotting, considering local threshold and global threshold illustrated in the Fig. 13(a). For local threshold, single query keyword was considered for optimization of the value, whereas a standard value has been used for all query keyword in case of global threshold. Tandem feature provided better performance than PHOG feature alone. Using global threshold, the average precision by Tandem-PHOG was found to be 84.6%.

Next, a comparison with popular "*Tesseract*" OCR is performed for printed text recognition. Here, the text from printed section is first separated from the handwritten part using previous step. Next, printed line images are fed into OCR which extracts and recognizes text (See Fig. 14). The ASCII text obtained from OCR is next searched to find the words of doctor's specialization/department list. Sometimes Tesseract may lead to partial recognition and then searching through a particular keyword creates a problem. To avoid this error, a string matching algorithm with edit distance has been integrated to predict the word present in the archived datasets.

It is noted that, using Tesseract, we achieved 74.62% of accuracy in our dataset with 50 keywords. The errors are generated mainly due to scanned images having low resolution. Sometimes due to slant and skew of word image, it produced error. After applying the edit distance based string matching some errors were corrected. During matching with a list of words, it predicts the correct word.

Query Keyword	Printed text line image	Result
EYE	ICLEP TRAINED AT. L.V. PRASAD EYE INST., HYDERABAD,	✓
EYE	MBBS., (Cal) DO (Cal), EYE Surgeon	✓
Dermatology	Assistant Professor of Dermatology	✓
Dermatology	M.B.B.S. (Cal) M.D. (Dermatology) DNB	✓
Dental	attached to : R. Ahmed Dental College and Hospital	✓
Dental	Consultant, Oral and Dental Surgeon	✓
Urological	CONSULTANT NEUROLOGIST	✗
Urological	Consultant Urological Surgeon	✓

Fig. 12. Qualitative result showing word spotting for printed text lines.

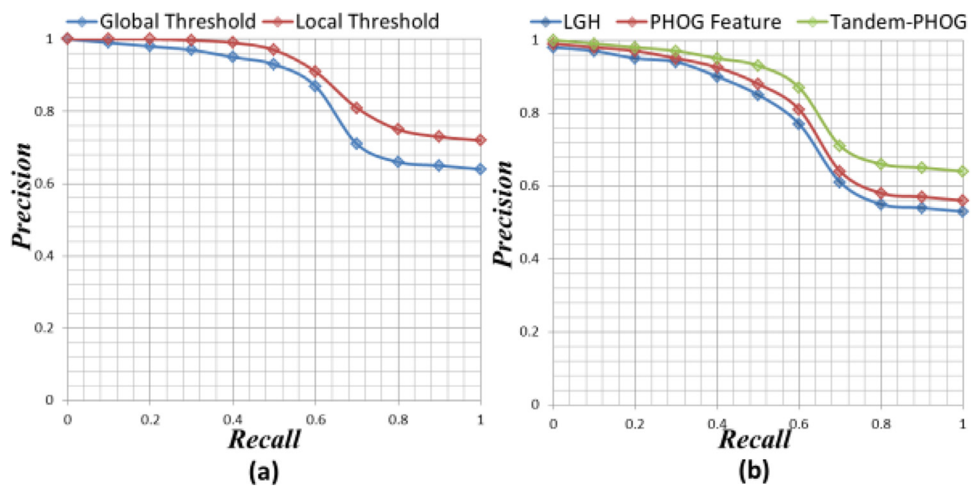

Fig. 13. Precision Recall curve (a) using Global and Local threshold using Tandem-PHOG Feature and (b) comparative study of PHOG and Tandem Feature for keyword spotting in printed text line.

Consultant Interventional Cardiologist	Consultant Interventional Cordlologist	Consultant Interventional Cardiologist
Professor of Paediatrics (Retd.)	Professor of Poediatrics (Retd.)	Professor of Paediatrics (Retd.)
CONSULTANT NEUROLOGIST	CONSULTANT KEUROLOGIST	CONSULTANT NEUROLOGIST

Fig. 14. Few examples of printed text and corresponding OCR output results. Note that, the error in some output (2nd column) is corrected by our system using string edit distance (3rd column).

5.3.1. Error analysis

One of these errors occurs in the image acquisition step where a blurred image may result in faulty recognition or even no recognition. Thus images were scanned taking special care so that noise level is eradicated. In prescriptions, doctors' handwriting has great variations and hence pre-processing was a complex process. Sometimes, the printed letter-head part contains text colour and textured background. As our present framework did not consider colour image processing, images having colour background will not be recognized by this system (See Fig. 15(a)).Also, we noted that the words which were abbreviated, like "Ortho" instead of "Orthopaedics" (Fig. 15(b)) were not detected using our system.

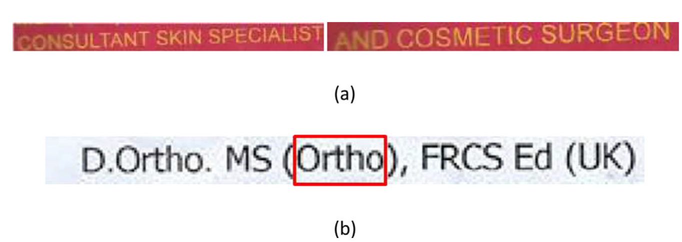

Fig. 15. Some examples of qualitative results where our system detected and failed to detect keywords. (a) Image with colour background. (b) Department name is abbreviated. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

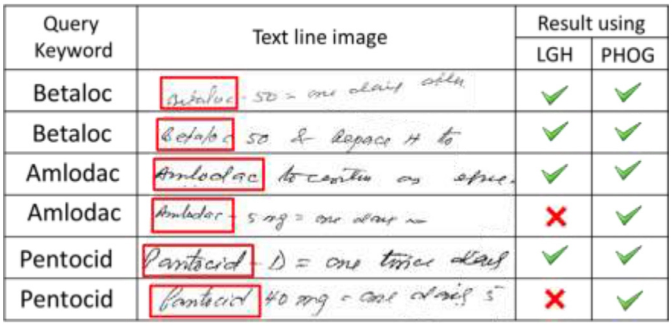

Fig. 16. Comparative study between LGH and PHOG feature for word spotting.

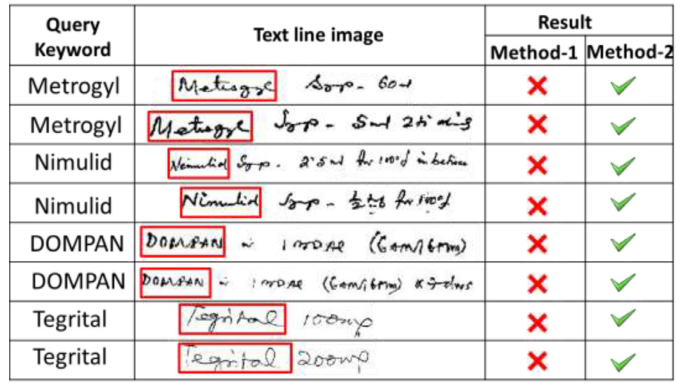

Fig. 18. Qualitative results where method-2(Using printed letter-head information) were found to give correct result but mehod-1(without using printed letterhead information) failed.

5.4. Performance evaluation of word spotting in handwritten text lines

To train our system in handwritten word spotting, the IAM (Marti & Bunke) English sentence dataset was used. The IAM data set consists of 1539 pages of handwritten modern English text written by 657 writers. We collected a total of 6161 hand written text line image from IAM dataset for training. For testing our approach a total of 1212 number of handwritten text line images were considered from the prescription image. A total of 140 keywords have been used in our experiment. These contain both disease and medicine names. Some examples of our word spotting mechanism in handwritten prescription using LGH and PHOG features are shown in Fig. 16. Note that, with keywords "Amlodac" and "Pentocid", LGH fails to detect the words. Comparative analysis of LGH, PHOG and Tandem-PHOG using Precision-Recall curve is shown in Fig. 17(b). Tandem-PHOG feature out performs LGH and PHOG feature in word spotting work. The experimental evaluation of Tandem-PHOG feature with global and local thresholds are presented in Fig. 17(a).

To compare our approach using traditional word matching method, DTW-based model (Rath & Manmatha, 2007) is used for similarity matching of handwritten words. The sequences are "warped" non-linearly in the time dimension to determine a measure of their similarity, independent of certain non-linear variations in the time dimension. A word is represented as 4 sequences, which are computed from four profiles features of a word. Four components such as vertical projection profile, upper profile, lower profile and vertical crossing (vertical ink transition) have been extracted as discussed in (Rath & Manmatha, 2007) for word-level features. The value of all the profile features is normalised to the range (0–1). The DTW-based technique for measuring similarity between two sequences uses the Sakoe-Chiba band (Sakoe & Chiba, 1978) to speed up the computation. The performance using DTW was found not satisfactory. We have obtained precision of 21.51% and recall of 17.14% respectively. The poor accuracy is mainly due to complex handwritten words in prescription.

5.5. Experiment with end-to-end system

It is to be noted that with knowledge information using department/discipline word from printed letterhead, the overall accuracy has been improved. The improvements in some text lines are shown in Fig. 18. Note that, department wise word spotting improves the overall performance. In Fig. 19 we have given the precision-recall curve to show the performance of our word-spotting frame work using tandem feature, where it has been observed that use of printed letterhead information gives an increase in the MAP value from 49.36 to 64.88. Also Fig. 19 shows that Tandem feature is found to be better than plain PHOG feature. We have obtained a maximum of 64.88 MAP value using Tandem-PHOG feature for global threshold with the information of department from the printed letter-head information (see Fig. 20). On the other hand, only 49.36% MAP was obtained without using printed letter head information.

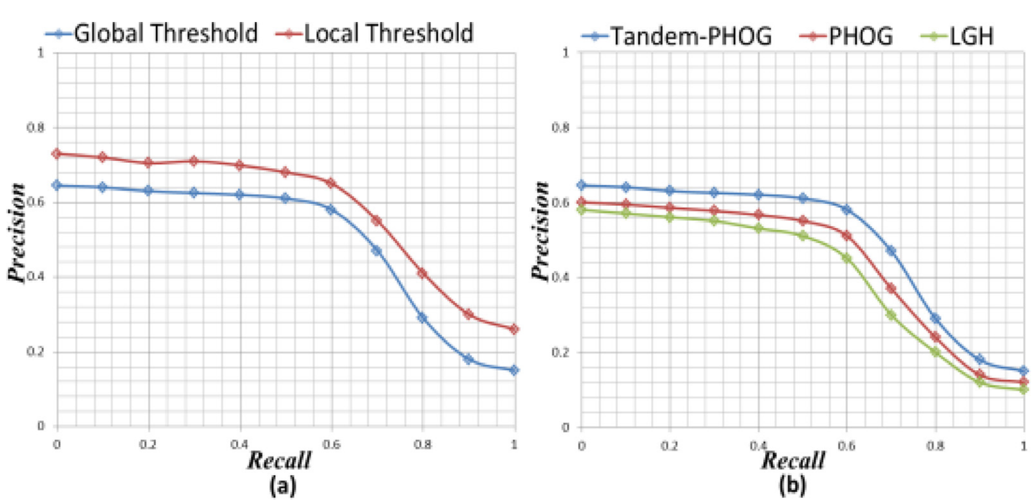

Fig. 17. Precision Recall curve (a) using Global and Local threshold using Tandem-PHOG Feature and (b) comparative study of Tandem-PHOG, PHOG and LGH Feature for keyword spotting in handwritten text line.

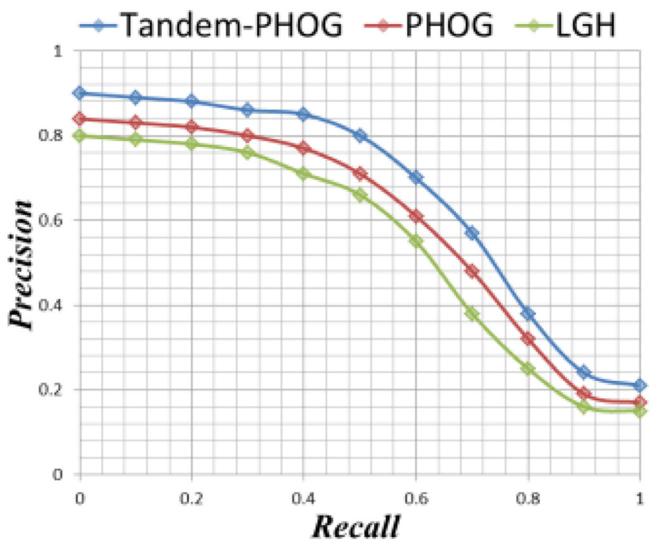

Fig. 19. Comparative study of Tandem-PHOG, PHOG and LGH Feature for keyword spotting in handwritten text line using information from printed letter-head.

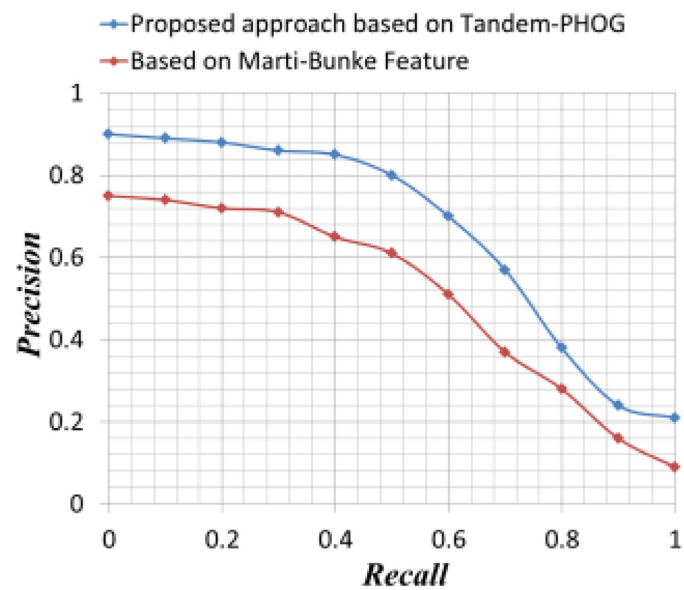

Fig. 21. Comparison with Marti-Bunke feature.

5.5.1. Comparison with other state-of-the-art feature

To compare the proposed PHOG feature, we considered the profile feature proposed by Marti and Bunke (2001) which have been used in Latin text recognition. It consists of nine features computed from foreground pixels in each image column. Three global features are used to capture the fraction of foreground pixels, the centre of gravity and the second order moment. Remaining six local features comprise of the position of the upper and lower profile, the number of foreground to background pixel transitions, the fraction of foreground pixels between the upper and lower profiles and the gradient of the upper and lower profile with respect to the previous column, which provides dynamic information. Fig. 21 shows the comparative study of our method with Marti-Bunke feature. Note that the proposed framework with Tandem-PHOG feature outperforms the Marti-Bunke features.

Table 3
MAP values using different number of Gaussians in HMM.

Printed letterhead information in handwritten word spotting	Feature	16 G	32 G	64 G
Not Used	LGH	45.51	46.14	45.94
	PHOG	46.23	47.35	46.68
	Tandem-PHOG	48.10	49.46	48.55
Used	LGH	59.94	61.57	60.34
	PHOG	61.25	63.14	61.58
	Tandem-PHOG	63.13	64.88	63.65

5.5.2. Parameter evaluation

We considered continuous density HMMs with diagonal covariance matrices of GMMs in each state. A number of Gaussian mixtures were tested on validation data. Word spotting performance with different Gaussian numbers is detailed in Table 3. It was observed that with 32 Gaussian Mixture, all features provide the best

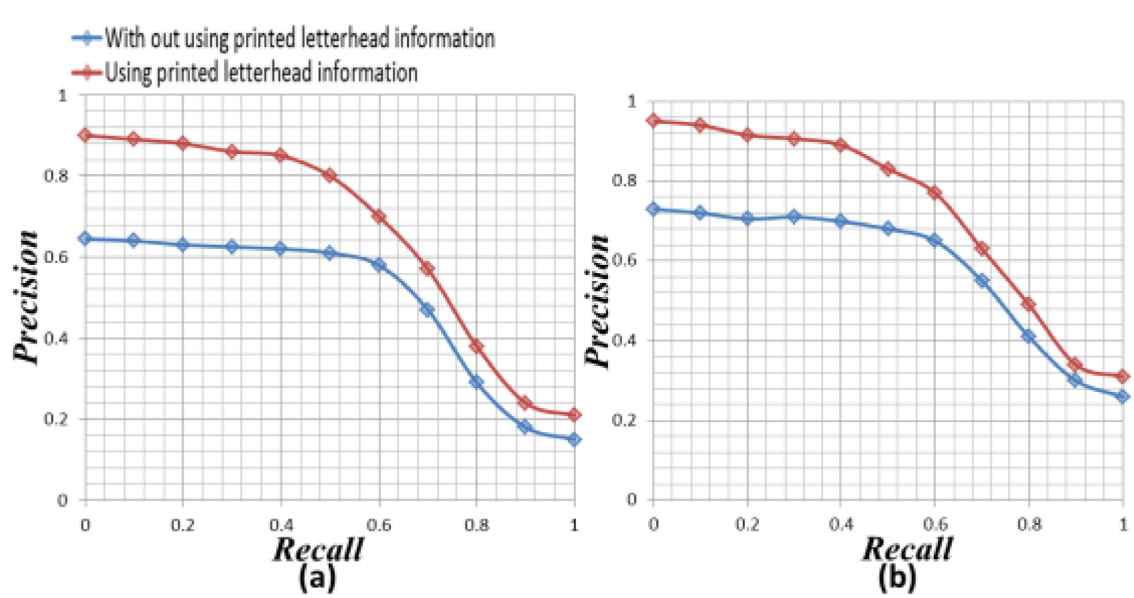

Fig. 20. Comparison between two methods namely word spotting using printed letterhead information and With-out using printed letterhead information. Global threshold has been considered for (a) and local threshold for (b).

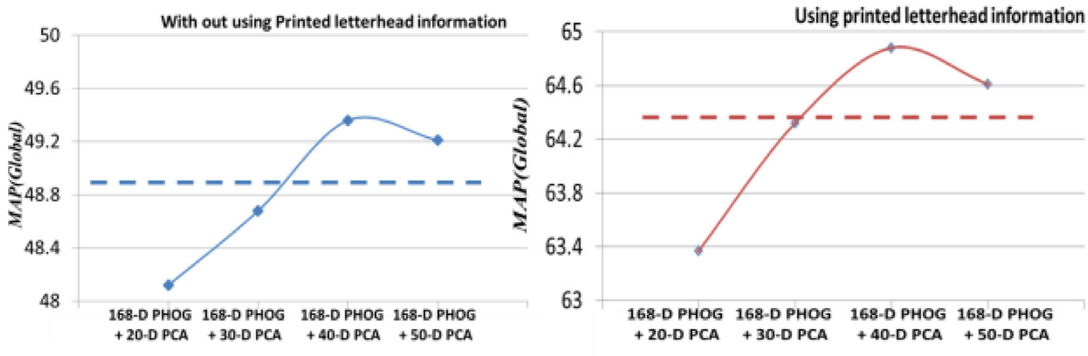

Fig. 22. Performance evaluation from PCA feature reduction.

Table. 4
Dataset description for different size of keywords.

Mode	Size of the Keywords							
	4	5	6	7	8	9	10	11
Printed	1.5%	5.1%	15.4%	20.3%	21.5%	16.1%	12.1%	7.7%
Handwritten	7.5%	11.4%	17.4%	19.2%	18.8%	16.7%	6.3%	2.6%

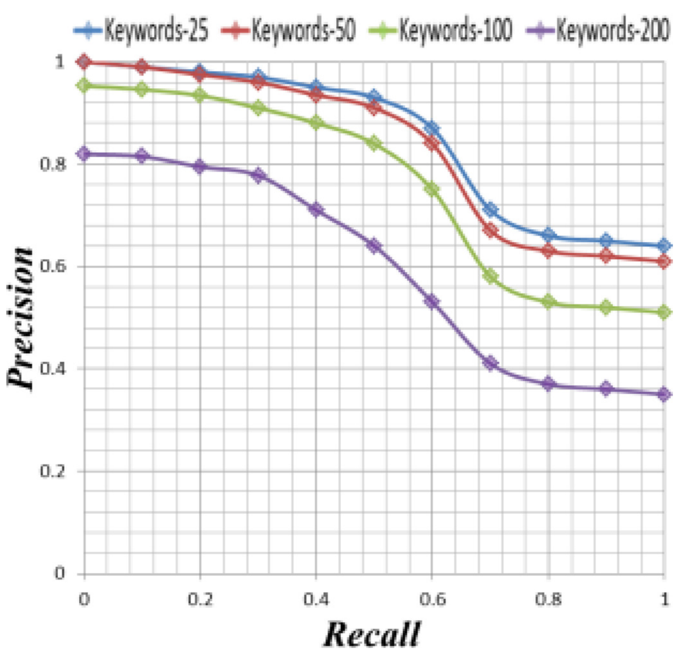

Fig. 23. Comparative study of word spotting performance using Tandem-PHOG in printed text line with different number of keywords.

results. The best state number was found to be 6. Fig. 22 illustrates the performance with varying the dimension of posterior feature on word spotting experiment. 64 hidden nodes were used in MLP during the experiment. With 40 features we obtained best accuracy in both word spotting approaches with and without letterhead information.

5.6. Scalability measurement

The scalability of our system is tested with increasing number of keywords in both printed and handwritten word spotting. For word spotting in printed text lines, the approach is tested with maximum 200 query words. In Table 4, the distribution of keywords according to their length is shown. Fig. 23 shows the precision-recall results with 25, 50, 100 and 200 keywords. With

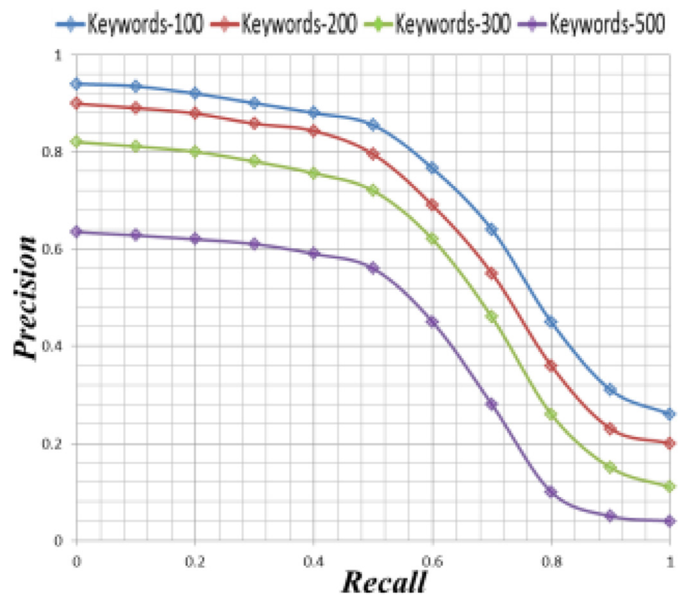

Fig. 24. Comparative study of word spotting performance in handwritten text line with different number of keywords using look-up table.

increasing number of keywords the precision degrades a little. Keeping recall to 0.6, the precision is higher than 75% with 100 keywords. Fig. 23 shows an analysis of the performance on the keyword length using our word spotting framework in printed text lines.

Similarly, word spotting in handwritten text lines is tested with a maximum 500 query words. In Table 4, the distribution of keywords according to their length is shown. Fig. 24 shows the precision-recall results with different number of keywords. Fig. 25 shows an analysis of the performance on the keyword length using our word spotting framework with printed letter head information.

5.7. Runtime evaluation

Experiments have been done on a computer with 2.80 GHz I5 and 4GB RAM with windows operating system using MATLAB. For each query, the average runtime has been computed from different runs made in the experiment. The time computation using different features are detailed in Table 5. The computation time Tandem-PHOG is more than other methods due to additional computation of character alignments and Tandem feature usage. Optimization and parallelization of the code are yet to be done, which promises better performances in terms of speed improvement. We have also checked the computation time for word retrieval using printed let-

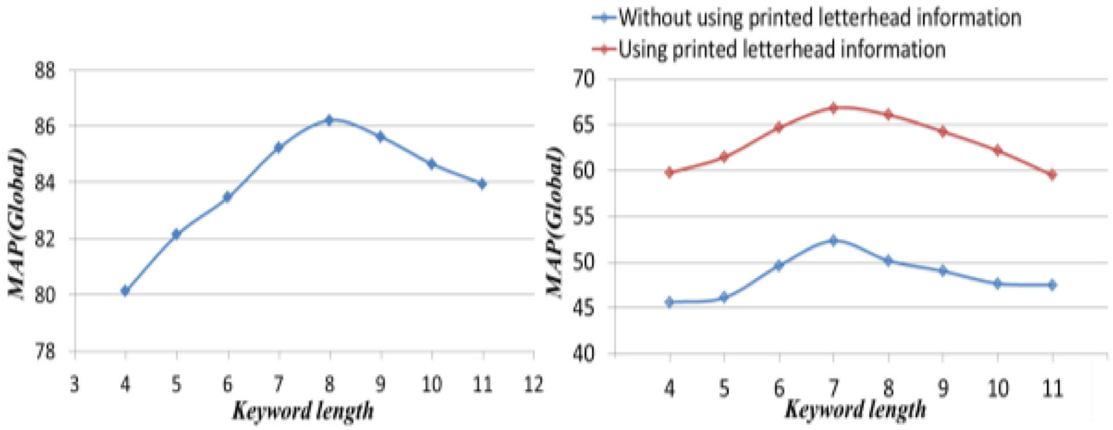

Fig. 25. Word spotting performance using keywords of different length for (a) Printed Part (b) Handwritten Part.

Table 5
Runtime comparison.

Word Spotting Approach	LGH	PHOG	Tandem-PHOG
Printed text line	1.29 Sec	1.38 Sec	1.64 Sec
Handwritten text line	1.43 Sec	1.53 Sec	1.78 Sec
Handwritten line using printed text information	0.72 Sec	0.78 Sec	0.97 sec

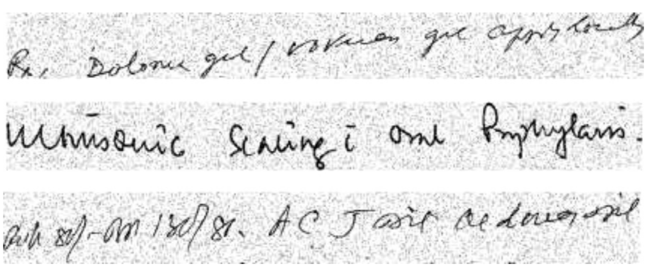

Fig. 26. Example of noisy image.

Table 6
MAP (Global) performance with synthetic noise.

Noise Parameters	Printed letterhead information	
	Used	Not used
10%	64.41	49.01
20%	62.47	47.58
30%	59.45	44.28

Table 7
Comparison of our framework in IAM (English) Dataset using global threshold. The mean average precision (MAP) is mentioned with the global threshold scenario.

Dataset	LGH	PHOG	Tandem-PHOG
IAM(English)	47.04	48.98	51.42

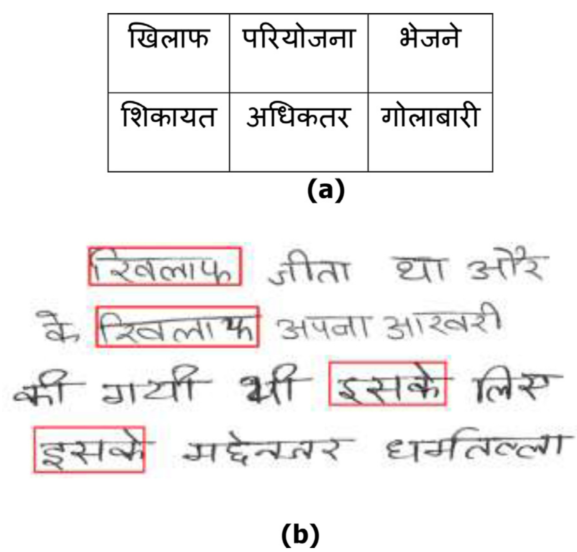

Fig. 27. Word spotting in Devanagari handwritten text lines (a) Query keywords. (b) Word spotting results in text line.

terhead information. It was found that, with printed text information integration the query search time reduces significantly.

5.8. Experiment with synthetic noise

We have tested our system with the lines by adding synthetic noises. The text lines are degraded with Gaussian noise of different noise levels (10%, 20% and 30%). For word spotting in such noisy text lines, our approach is as follows. Tandem-PHOG features are extracted from each text line using sliding window. Next, HMM-based word spotting technique is applied to find the keywords using the sliding window features. To get an idea of word spotting results, some line images are added with 20% Gaussian noise and corresponding results are shown in Fig. 26. Quantitative results with noisy images obtained by different level of Gaussian noise are shown in Table 6.

5.9. Evaluation of feature on other dataset

5.9.1. Experiment on IAM dataset

To compare the performance of Tandem-PHOG feature, we have used IAM dataset for evaluation. The testing dataset of IAM English sentence dataset (Marti & Bunke, 2002) has been used to evaluate the performance. Total 929 text lines and 450 keywords were considered for comparison. Next, LGH, PHOG and Tandem features are extracted from these text images and HMM based word spotting is applied for word spotting. To summarize the performance of a system with a single figure, the average precisions are measured for the 450 keywords and the mean is reported. Table 7 shows the MAP with best performance obtained using Tandem-PHOG features.

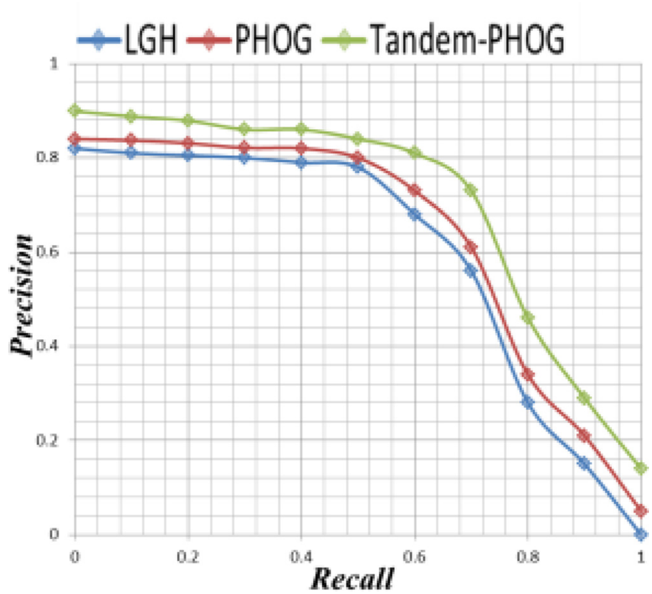

Fig. 28. Comparative study between LGH, PHOG and Tandem-PHOG feature for word spotting Devanagari handwritten text lines.

5.9.2. *Experiment on competition dataset*

Recently, two competitions on handwritten keyword spotting have been organized which put light on the recent state of the arts of keyword spotting in the literature. For a fair comparison with our proposed method of word spotting, we have tabularized the performance accuracies (in terms of % mAP) of the top entries of ICDAR'15 (Puigcerver, Toselli, & Vidal, 2015) and ICFHR' 14 (Pratikakis, Zagoris, Gatos, Louloudis, & Stamatopoulos, 2014) handwritten word spotting competitions along with our method. The datasets in ICDAR'2015 competition consist of a series of handwritten documents prepared by tranScriptorium project. From 70 original handwritten pages of different handwritings, styles and fonts, a total of 15,419 word images with proper annotations have been used. Some of these approaches used segmentation free methods and applied Query-by-Example (QbE) to search keywords. The approach, "Argus" used Query-by-String (QbE) to search keywords in segmented lines. Their recurrent Neural Network (RNN) based approach showed better results than our performance. The improvement using RNN is mainly due to the discriminative training in neural network.

In the competition of ICFHR'14, two datasets have been used namely 'Bentham' (consists of high quality handwritten manuscripts written by Jeremy Bentham and colleague) and 'Modern' (consists of modern document written by several writers and in several languages). It is observed from the table that systems without segmentation perform poorly than systems with segmentation (either line or word level). It is also noted that our Tandem-

Table 8
Comparative study with recent handwritten word spotting competition.

ICDAR '15 Competition on Handwritten Keyword Spotting

Algorithm	Main Method	Configuration	Accuracy
ARGUS, jointly developed by CITlab and PLANET intelligent systems, GmbH	Recurrent Neural Network (RNN)	Segmentation: Line Training: Yes Type: QbS	**mAP: 87.11%**
A feature-descriptor based system by PRG, TU Dortmund University, Germany	Local SIFT descriptor + Bray-Curtis distance for Ranking	Segmentation: Word Training: No Type: QbE	mAP: 42.44%
BoVW framework by CVC, Universitat Autònoma de Barcelona, Spain	Bag of Visual Words framework	Segmentation: Page Training: No Type: QbE	mAP: 8.21%
Our Method	**Tandem HMM model**	**Segmentation: Line** **Training: Yes** **Type: QbS**	mAP: 84.91%

ICFHR '14 Competition on Handwritten Keyword Spotting

Algorithm	Main Method	Configuration	Accuracy On Bentham	Accuracy On Modern
HOG & LBP descriptor based system by The Blavatnik School of Computer Science, Tel-Aviv University, Israel	k-NN search based retrieval using 250D descriptor	Segmentation: Page Training: No Type: QbE	**mAP: 52.4%**	mAP: 33.8%
Semantic representation learning based system by CVC, *Universitat Autònoma de Barcelona, Spain*	PHOC descriptor	Segmentation: Word Training: Yes Type: QbE	mAP: 51.3%	mAP: 52.3%
HOG & LBP descriptor based system by The Blavatnik School of Computer Science, Tel-Aviv University, Israel	k-NN search based retrieval using 250D descriptor	Segmentation: Page Training: No Type: QbE	mAP: 41.6%	mAP: 26.3%
One-shot handwritten word spotting by Smith College, Dept. Of CS, Northampton MA, USA	Flexible Inkball model	Segmetation: Page Training: Yes Type: QbE	mAP: 36.3%	mAP: 16.3%
Our Method	**Tandem HMM model**	**Segmentation: Line** **Training: Yes** **Type: QbS**	mAP: 50.12%	**mAP:51.01%**

HMM based word spotting methods increase the performance by a significant margin.

5.9.3. Experiment on Indic dataset

To test the robustness of Tandem-PHOG features, we have considered handwritten documents of Indic (Devanagari) script for our performance evaluation. We have collected word images of different writer for both Bangla and Devanagari script. Then we have generated a total of 8592 line images for Devanagari containing two to six word images in a line. We considered 810 line images as validation, 878 as testing and rest as training. A total of 205 keywords were taken for our experiment. Some examples of qualitative results are shown in Fig. 27. For quantitative measures, an average of the results is taken. Comparisons with LGH, PHOG and Tandem-PHOG features have been performed. Fig. 28 shows the comparative study of these features using precision and recall. It was noted that PHOG-Tandem feature outperforms other features in word spotting.

6. Conclusion

In this paper, a keyword spotting based approach is proposed for information retrieval from handwritten medical prescription images. It includes a novel retrieval technique of medicine/disease names in the prescription archive. Our contribution in this paper is two-folded. First, a novel Tandem-HMM system has been proposed for word spotting in handwritten text documents. From the experiment study with existing approaches we have shown that Tandem system provides better performance. We also provided comparative results in Table 8 in competition datasets to highlight the improvement than existing approaches. Also, we demonstrated how domain knowledge can be useful to improve the traditional word detection approaches. The domain knowledge has been used to reduce the target text information which in return improves the performance significantly. The experiments show that the keyword retrieval accuracy in handwritten documents has been improved by 15.42% using the knowledge extracted from printed letterhead. The experiment is also performed by adding synthetic noises to original prescription images to provide robustness of our approach. Since, the proposed approach uses knowledge from printed text information, if the printed-text information is not retrieved properly from prescription letterhead due to noise and not-common fonts, the keyword spotting in handwritten documents may not be effective.

The proposed system attempts to eradicate the problems that a patient faces while reading a prescription because of poor handwriting of doctors. Moreover, many hospitals recently have started maintaining a printed prescription dataset which is quite useful while retrieving the medical history of any patient and while referencing a patient from one hospital to another. Our system will allow to keep the records of handwritten medical prescriptions which may be used for various purposes like, academic purposes, i.e. while studying the case history of a patient, in legal purposes and for research works. It can keep a record of various medicines prescribed by doctors in different period of time to understand the usefulness of the medicines. Also, our proposed system can be developed to detect wrong medication if it happens unintentionally. In future we would like to progress for automatic transcription of such prescriptions.

Acknowledgements

The authors would like to acknowledge the support of A. Halder and A.Mukherjee of the Dept. of CSE, Meghnad Saha Institute of Technology, Kolkata, India, towards the development of the database described in the present article.

References

Almazán, J., Gordo, A., Fornés, A., & Valveny, E. (2014a). Word spotting and recognition with embedded attributes. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 36*(12), 2552–2566.

Almazán, J., Gordo, A., Fornés, A., & Valveny, E. (2014b). Segmentation-free word spotting with exemplar SVMs. *Pattern Recognition, 47*(12), 3967–3978.

Antonacopoulos, A., & Downton, A. C. (2007). Special issue on the analysis of historical documents. *International Journal on Document Analysis and Recognition*, 75–77.

Bhunia, A. K., Das, A., Roy, P. P., & Pal, U. (2015). A comparative study of features for handwritten bangla text recognition. In *Proceedings of the ICDAR* (pp. 636–640).

Bishop, C. M. (2009). *Pattern recognition and machine learning*. Springer 2006.

Bunke, H. (2003). Recognition of cursive roman handwriting: Past, present and future. In *Proceedins of the ICDAR* (pp. 448–459). 2003.

Cao, H., & Govindaraju, V. (2007). Template-free word spotting in low-quality manuscripts. In *Proceedings of international conference on advances in pattern recognition* (pp. 135–139).

Cao, H., Prasad, R., & Natarajan, P. (2011). Handwritten and typewritten text identification and recognition using hidden Markov models. In *Proceedings of the ICDAR* (pp. 744–748).

Chen, Q., Gong, T., Li, L., Tan, C. L., & Pang, B. C. (2010). A medical knowledge based postprocessing approach for doctor's handwriting recognition.. In *Proceedings of the international conference on frontiers in handwriting recognition* (pp. 45–50).

Fischer, A., Keller, A., Frinken, V., & Bunke, H. (2012). Lexicon-free handwritten word spotting using character HMMs. *Pattern Recognition Letters, 33*(7), 934–942.

Frinken, V., Fischer, A., Manmatha, R., & Bunke, H. (2012). A novel word spotting method based on recurrent neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 34*, 211–224.

Guo, J. K., & Ma, M. Y. (2001). Separating handwritten material from machine printed text using hidden markov models. In *Proceedings of the ICDAR* (pp. 439–443).

Hermansky, H., Ellis, D. W., & Sharma, S. (2000). Tandem connectionist feature extraction for conventional HMM systems. In *Proceedings of IEEE international conference on acoustics, speech, and signal processing, (ICASSP): 3* (pp. 1635–1638).

Jayadevan, R., Kolhe, S. R., Patil, P. M., & Pal, U. (2012). Automatic processing of handwritten bank cheque images: A survey. *International Journal on Document Analysis and Recognition (IJDAR), 15*, 267–296.

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

Liu, Cheng-Lin, Koga, Masashi, & Fujisawa, Hiromichi (2002). Lexicon-driven segmentation and recognition of handwritten character strings for Japanese address reading. *IEEE Transactions on Pattern Analysis and Machine Intelligene, 24*, 1425–1437.

Marti, U. V., & Bunke, H. (2002). The IAM-database: An English sentence database for offline handwriting recognition. *International Journal on Document Analysis and Recognition, 5*, 39–46.

Milewski, R. J., Govindaraju, V., & Bhardwaj, A. (2009). Automatic recognition of handwritten medical forms for search engines. *International Journal of Document Analysis and Recognition (IJDAR), 11*, 203–218.

Nagy, G., & Lopresti, D. (2006). Interactive document processing and digital libraries. In *2nd international workshop on document image analysis for libraries* (pp. 2–11).

Niyogi, D., Srihari, S. N., & Govindaraju, V. (1996). Analysis of printed forms. *Handbook on optical character recognition and document image analysis*.

Pratikakis, I., Zagoris, K., Gatos, B., Louloudis, G., & Stamatopoulos, N. (2014). ICFHR 2014 competition on handwritten keyword spotting (H-KWS 2014). In *Proceedings of the international conference on frontiers in handwriting recognition* (pp. 814–819).

Puigcerver, J., Toselli, A. H., & Vidal, E. (2015). ICDAR2015 competition on keyword spotting for handwritten documents. In *Proceedings of the ICDAR* (pp. 1176–1180).

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

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

Rodríguez-Serrano, J. A., & Perronnin, F. (2012). A model-based sequence similarity with application to handwritten word spotting. *IEEE transactions on pattern analysis and machine intelligence, 34*(11), 2108–2120.

Rose, R. C., & Paul, D. B. (1990). A Hidden Markov Model based keyword recognition system. In *Proceedings of the international conference on acoustics, speech, and signal processing* (pp. 129–132).

Rothfeder, J. L., Feng, S., & Rath, T. M. (2003). Using corner feature correspondences to rank word images by similarity. In *Proceedings of the computer vision and pattern recognition workshop* 30-30.

Roy, P. P., Pal, U., & Lladós, J. (2008). Morphology based handwritten line segmentation using foreground and background information. In *Proceedings of the international conference on frontiers in handwriting recognition* (pp. 241–246).

Roy, P. P., Ramel, J. Y., & Ragot, N. (2011). Word retrieval in historical document using character-primitives. In *Proceedings of the ICDAR* (pp. 678–682).

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

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. In *IEEE transactions onacoustics, speech and signal processing: 26* (pp. 43–49).
Srihari, S. N., & Kuebert, E. J. (1997). Integration of hand-written address interpretation technology into the united states postal service remote computer reader system. In *Proceedings of the ICDAR* (pp. 892–896).
Srihari, S. N., Huang, C., & Srinivasan, H. (2005). Search engine for handwritten documents. Electronic Imaging 2005. *International Society for Optics and Photonics*, 66–75.
Wshah, S., Kumar, G., & Govindaraju, V. (2014). Statistical script independent word spotting in offline handwritten documents. *Pattern Recognition, 47*, 1039–1050.
Zhang, B., Srihari, S. N., & Huang, C. (2003). Word image retrieval using binary features. Electronic imaging 2004. *International Society for Optics and Photonics*, 45–53 December.
Bluche, T., Ney, H., & Kermorvant, C. (2013). Feature extraction with convolutional neural networks for handwritten word recognition. In *Proceedings of the ICDAR* (pp. 285–289).
Marti, U. V., & Bunke, H. (2001). Using a statistical language model to improve the performance of an HMM-based cursive handwriting recognition system. *International Journal on Pattern Recognition and Artificial Intelligence, 15*, 65–90.