www.ijecs.in

International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 4 Issue 6 June 2015, Page No. 12952-12955

Recognition Text for Liver Cell Lab Reports Based on Hybrid HMM Approach

K.Lokanayaki[1], *Dr.A.Malathi*[2]

[1]Department of Computer Application, Spurthy Group of Institutions,
Bangalore, Karnataka, India
kloganayakischolar@gmail.com

[2]Assistant Professor, PG and Research, Department of Computer Science
Government Arts College, Coimbatore-18
malathi.arunachalam@yahoo.com

Abstract: ***A Form-based Optical Word Recognition (OWR) System for printed forms includes functional mechanism for extract word from image and word image recognition. The recognizer is a very important component of the OWR system. Automatic recognition for printed images is a complex task. In this paper we discuss the form image recognition technique and improved field matching techniques implemented in our system. These effective techniques help in preparing the input word image for the Hidden Markov Model (HMM) based recognizer and Partial Matching (PM) algorithms for improving overall recognize matching accuracy. Although these algorithms have been implemented with Liver Cell reports in our OWR.***

Keywords: Form-based OWR, Word image extraction, Hidden Markov Model, Partial Matching**.**

1. Introduction

In biomedical images in computer-based patient report systems can become increasingly difficult to maintain and retrieve the information. Here also the fixed textual data such as patient reports or patient lab reports are being filled with X-ray images, MRI scans, CT scans, and video streams.

Efficient filtering for digital medical images.In this paper, we present a report-based matching algorithm that can recognize textual information from medical report images. It is used to covert pixel information to text. Finally, text will be stored to database.

Data mining algorithms described here were recently used in many biomedical applications. Sural& Das [1] present a concept of fuzzy sets for recognizing Bangla script. This technique has defined the Hough transform of character pattern pixels from which additional fuzzy sets are synthesized using t-norms.It also trained with a more number of linguistic set memberships derived from these t-norms.

Fuzzy Logic [6] is also used for Matrix Matching converts each character into a pattern within a matrix and then compares the pattern with an index of known characters. Its recognition is strongest on monotype and uniform single column pages. It is a multi-valued logic that allows intermediate values to be defined between conventional evaluations like yes/no, true/false, black/ white etc.

An attempt is made to attribute a more human-like way of logical thinking in the programming of computers. Fuzzy logic is used when answers do not have a distinct true Mahmud et al [2] also taken Bangla multi font characters recognized isolate and continuous printed characters segmentation.

In [3] feed-forward neural network used for classification of recognition data.In this strategy they simulate the Neural Networks. This technique taken all pixels of in each image and matches them to a known indexof character pixel patterns.

The ability to recognize characters through abstraction is great for faxeddocuments and damaged text. It also found specific types of problems, such as processing stockmarket data or finding trends in graphical patterns.

2. Related Work

During the last decade, text extraction has become the biggestissue in research field. In this recognition, extract text from automatic speech recognition [9], handwriting recognition [7] and printed text [12].

Majumdar [8] has proposed based on K-nearest neighbour classifier using digital curvelet transform for recognizing text from Bangal Multi-font basic character images.It also used for feature extraction from an original image as well as its morphologically altered versionsare used to train a set of *K*-nearest neighbour classifiers.

The output value of these classifiersare fused using a simple majority voting scheme to arrive at a final decision. Apriori algorithm [10, 11] used for elucidate relationships in data was frequent item sets used to association rules.This algorithm used for commonly applied to transactional dataset.

This algorithm assembles frequent item sets into association rules that indicate conditional frequency. The [17] Morphological Component Analysis (MCA) algorithm used for successful text extraction from graphical document images based.

It is also used for sparse representation framework with two appropriately chosen discriminative over complete dictionaries. This method overcomes the problem of touching between textand graphics and also insensitive to changed font sizes, styles and orientations.

Haar discrete wavelet transform also used to detect edges of candidate text regions [18]formulated an efficient and computationally fast method from documents. In this methodmainly used to morphological dilation operator to connect the isolated candidate text edge and then a linefeature vector graph was generated based on the edge map.

2.1 Hidden Markov Models (HMMs)

This method recognized for continuous speech, printed or handwritten and cursive script text recognition tasks[13], [14], [15].HMMs are statistical models in which system being modelled is considered as a Markov process that have unobserved or hidden states[16].

In a Hidden Markov Model, state is not directly visible but it is associated with a probability distribution over all possible output values. Each state is associated to an input pattern and is modelled by a probability distribution function .Random process: random memory less walk on a graph of nodes called states.

Parameters:

- Set of states $S = \{S_1, ..., S_n\}$ that form the nodes
- Let q_t denote the state that the system is in at time t
- Transition probabilities between states that form the edges, $a_{ij}= P(q_t = S_j \mid q_{t-1} = S_i), 1 \le i,j \le n$
- Initial state probabilities, $\pi_i = P(q_1 = S_i), 1 \le i \le n$

2.2 Partial Matching Algorithm

In[19] partial-match query q with s keys specified.It represented by a record r= R with k- s keys replaced by the special symbol "*" (meaning "unspecified"). If f= (f_1.... f_n), then for k-s.The set q() is the set of all records agreeing with inthe specified positions. Thus

$q(\Sigma^k)= \{ r\varepsilon\Sigma^k|[(\forall j, 1<=j<= k)[(ri==*)V(r1==ri)]\}$.

A sample application might be a crossword puzzle dictionary, where a typical query could require finding all words of the form "B*T**R" (that is: BATHER, BATI'ER, BETTER, BETTOR, BITTER, BOTHER, BUTLER, BUTTER).

We shall use Os throughout to denote the set of all partial-match queries with s keys specified.

3. Proposed work:

After document binarization a top-downsegmentation approach is applied. First lines of thedocuments are detected, then words are extracted andfinally match word and label word, the index value stored to database..

```
Step 1: Read the Initial values of image file, it will give the starting point of the file and the size of file.
Step 2: Repeat Step 3 to step while (Not end of Input file)
Step 3: Calculate Vertical Boundary value
Step 4: Calculate Horizontal Boundary value
Step 5: Get Line coordinates from Vertical and Horizontal value.
Step 6: Repeat Step 7 to Step While until Line coordinates not equal to NULL
Step 7: Calculate Vertical Boundary value of word
Step 8: Calculate Horizontal Boundary value of word
Step 9: Get window coordinates from Vertical and Horizontal value of word
Step 10: Repeat step 11 to step 13 until while word not equal to Null
Step 11: If word equal to label then
Step 12:  text ← word[index].
Step 13: else got to step 5
Step 14: go to step 2
```

This search process requires a word models, a possible word lexicon or dictionary, and a statistical language model [20].

The choice of lexicon and language model is optional. In presented system, we employ by building HMM with Partial (HMMP) matching algorithm for word level.

4. Experimental Results

We implemented our proposed word recognition and matching method for liver cell pathology report images. We collected these images from various hospitals for our experiments.

It consist various font name, font size, text, numbers, operators like +,-,etc and different distance. To evaluate the performance of proposed algorithm, we compare it with the well-know recognition based method like HMM [8] method and partial matching method.

The results of these methods were compared with the results achieved by our proposed method. All these methods have been evaluated based on window size.All the implementations of the proposed mod are done Matlab.

4.1 Recognition Accuracy

The proportion evaluation is carried out by computingword recognition accuracy percentage (WRA %) with the help of following formula

$$\text{WRA\%} = \frac{\text{N}-\text{ED}}{\text{N}} * 100$$

where N = Total number of words andED = Edit Distance = Nos. of deletions + Nos. of insertions + Nos. of substitutions (with equal cost).

4.2 **Error rate.**

Misclassified characters go by undetected by the system, and manual inspection of the recognized text is necessaryto detect and correct these errors. A low error rate may lead to a higher rejection rate and a lower recognition rate.

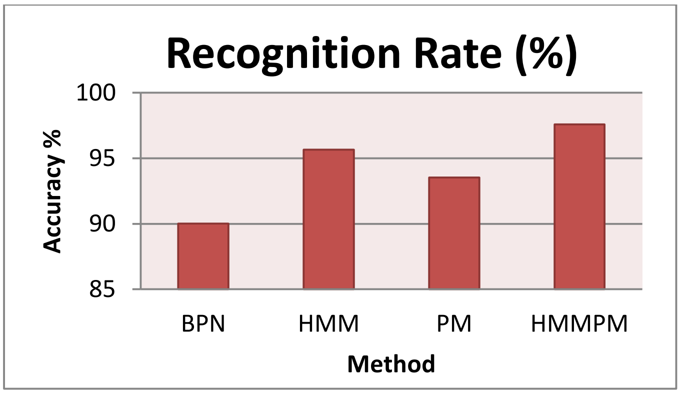

Figure 1. Detection Rate for Proposed Method

5. Conclusion

In this work we assess the hybrid HMM with matching technique for recognitionwords from liver cell report and store the each word to data base using partial matching algorithm. In these methods achieve more accurate recognition, low learning time and also more matching accuracy.

In our experiments of hybrid HMM with partial matching word recognition technique obtain better results for words database. A correct word recognition accuracy of proposed technique 97.58 %compare than existing technique.

References

[1] S Sural, PK Das "An MLP using Hough transform based fuzzy feature extraction for Bengali script recognition",Patter Recognition Letters 771-782.

[2] J.U. Mahmud, M.F. Raihan and C.M. Rahman, "A Complete OCR System for continuous Bengali Character", TENCON 2003, Conference on Convergent Technologies for Asia-Pacific Region, 15-17 Oct. 2003.

[3] A.O.M. Asaduzzaman et al., "Printed bangla textrecognition using artificial neural network with heuristicmethod," Proceedings of International Conference onComputer and Information Technology, 2002, pp. 27-28.

[4] Charikar M, Chen K, Farach-Colton M: Finding Frequent Items in Data Streams. In Proceedings of the 29th International Colloquium on Automata,Languages, and Programming ICALP. 2002:693–703.

[5] Park J, Chen M, Yu P: An effective hash-based algorithm for mining association rules. In Proceedings of the 1995 ACM SIGMOID international conference on Management of data ACM SIGMOID. 1995:175–186.

[6] Naulaerts S, Meysman P, Bittremieux W, Vu T, Berghe W, Goethals B, Laukens K: A primer to frequent itemset mining for bioinformatics.Briefings in Bioinformatics 2013, bbt074.

[7] Abuhaiba, S.Mahmoud, and R.Green, "Recognition ofHandwritten Cursive Arabic Characters*", IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol. 16, No. 6, June, 1994.

[8] MajumdarAngshul,"Bangla Basic Character Recognition Using Digital Curvelet Transform",Journal of Pattern RecognitionResearch, Vol 2, No 1 (2007).

[9] J. Makhoul, C. LáPré, C. Raphael, R. Schwartz, and Y.Zahao, "Towards language-independent character recognitionusing speech recognition methods," *in The 5th InternationalConference and Exhibition on Multi-Lingual Computing,Cambridge University Press*,1996.

[10] Christos Nikolaos E. Anagnostopoulos, Member, IEEE, Ioannis E. Anagnostopoulos, Member, IEEE, VassiliLoumos, Member, IEEE, and EleftheriosKayafas, Member, IEEE ,"A License Plate-Recognition Algorithm for Intelligent Transportation System Applications",ieee transactions on intelligent transportation systems, vol. 7, no. 3, september 2006, 377.

[11] Nafiz Arica and Fatos T. Yarman-Vural"An Overview of Character Recognition Focused on Off-Line Handwriting"ieee transactions on systems, man, and cybernetics—part c: applications and reviews, vol. 31, no. 2, may 2001.

[12] S.V. Rice, 1. Kanai, T.A. Nartker, "An technique toEvaluation of OCR Accuracy,"*Information Science Research Institute, 1993Annual Research Report,* University data. of Nevada, Las Vegas, pp. 9-20,1993.

[13] F. Jelinek, Statistical methods for speech recognition. Cambridge,MA, USA: MIT Press, 1997.

[14] I. Bazzi, R. Schwartz, and J. Makhoul, "An omnifontopenvocabularyOCR system for English and Arabic," IEEE Trans.Pattern Anal.Mach. Intell., vol. 21, pp. 495–504, June 1999.

[15] U.-V. Marti and H. Bunke, Using a statistical languagemodel to improve the performance of an HMM-based cursivehandwriting recognition systems. River Edge, NJ, USA:World Scientific Publishing Co., Inc., 2002, pp. 65–90.

[16] L. R. Rabiner, “Readings in speech recognition”, A. Waibeland K.-F. Lee, Eds., 1990, ch. A tutorial on Hidden MarkovModels and selected applications in speech recognition, pp.267–296.

[17] Thai V. Hoang , S. Tabbone(2010),“Text Extraction From Graphical Document Images Using Sparse Representation”in Proc. Das, pp 143–150

[18] Audithan,,R.M.Chandrasekaran (2009), "Document Text Extraction From Document Images Using Haar Discrete Wavelet Transform",European Journal Of Scientific Research, Vol.36 No.4 , pp.502-512.

[19] L.Ronald ,Rivest ,”Partial Matching Retrieval Algorithms”,SIAM J. COMPUT.Vol. 5, No. 1, March 1976.

[20] Victor Marti,Horst Bunke“Using a Statistical Language Model to Improve the Performance of an HMM-Based Cursive Handwriting Recognition System”International Journal of Pattern Recognition and Artificial Intelligence , Aug-2001, 15(01),65-90.