International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 04 | Apr-2016 www.irjet.net p-ISSN: 2395-0072

Word Segmentation Method for Handwritten Documents based on Structured Learning

Prof.K.N.Shedge, Jyoti Lokhande, Tushar jawale, Rahule Devare

M-Tech (cse), S.V.I.T, Chincholi, Nasik, Maharashtra – 422102.

B.E (Computer), S.V.I.T, Chincholi , Nasik, Maharashtra – 422102.

B.E (Computer) , S.V.I.T, Chincholi , Nasik , Maharashtra – 422102.

B.E (Computer) , S.V.I.T, Chincholi , Nasik , Maharashtra – 422102.

---***---

Abstract - ***In this paper we find out the writer with help of handwriting text document .It is very difficult task because of features of handwritten document based on variation and dissimilarity depend on person. so we have to proposed a new system using algorithm SIFT to identify person based on handwriting.***

Key Words **: Handwritten text image, MYSQL Database , structured SVM.**

1.INTRODUCTION:

Segmentation of handwritten document into its basic entities like line, words, curve is consider to be a crucial task in the field of handwriting document identification. different types of task are segmentation of line and segmentation of word procedure. the using SIFT based proposed algorithm is help us to solved the problem of writer identification to use of minimum steps and time to find out writer.

1.1 Proposed System:

The proposed method learns the local spatial relationship between text-lines and applies an effective strategy to predict the graph structure for partitioning. For the word segmentation document images are first segmented into text-lines, Then, the word segmentation algorithm (for a single text-line) is applied to individual text-lines. Although the characteristics of inter-word gaps are changing across (and even in) documents, there are strong correlations (e.g., scale) between them in a text-line.

1.2 Existing system

handwritten text recognition, larg contributed works to the related issues of classification of character, over transformation, model of language, evaluation of path and search, and estimation of parameter. For over segmentation, connected component analysis has been adopted widely, After candidate generating character patterns by combining consecutive primitive segments, each candidate pattern is classified using a classifier to assign match-unmatch scores to some character classes classification of character involves normalization of character, feature extraction, and design of classifier. Chinese characters classification of classes with large number, the most popularly used classifiers are the modified quadratic Discriminate function (MQDF) and the nearest prototype classifier (NPC).

2. SYSTEM DESIGN

In this system we are implement some modules for
User.
2.1Module
1. Writer Registration and Login:

Every writer first register her information name, mobile no, password, repassword, email add. Then all information of writer are save. Then writer login in this system.

2. Training the system:

In second module we train the system. we follow different steps like ostus binary image, connected component, isotropic LOG filter, threshold obtain ostus , remove blur and finally stored feature of text image document into a database.

3. Testing the document:

After the training we focused on testing. in this phase matching of feature of input text of database and extracted output text are match. The SIFT algorithm is used to match the feature then writer are identified.

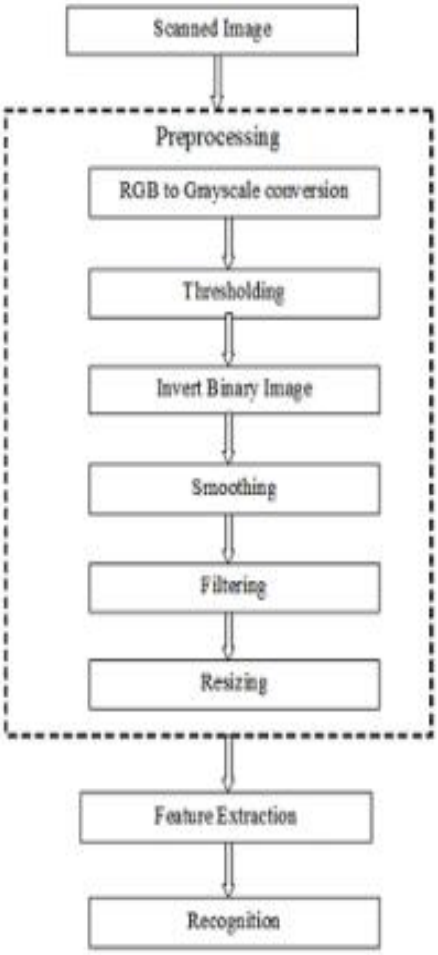

Fig -1: System Architecture

3. CONCLUSIONS

problem as a binary quadratic programming and estimated the parameters with the structured learning method. Due to the proposed formulation, we could take into account the pair wise similarities between word-separators as well as unary properties in the word segmentation. Also, due to the Structured SVM, all parameters are estimated in a principled way and it is believed that our method can be easily extended to other databases.

ACKNOWLEDGEMENT

We hereby take this opportunity to record our sincere thanks and heartily gratitude to our project guide Prof.K.N.Shedge for his useful guidance and making us available his intimate knowledge and experience in preparation of our project Word Segmentation Method for Handwritten Document based on Structure Learning. Sincere thanks to him for is valuable and time to time guidance and for rectifying all our errors, to make our seminar more accurate.

We are thankful to HOD of our Computer Department Prof. S. M. Rokade for his constant enlightenment and motivation which has been highly instrument al in making our seminar report.

The acknowledge will be incomplete if we do not record our sense of gratitude to our principal Prof. G. B. Shinde who gave us necessary guidance and encouraged by providing us with all the facilities available to work on this paper. Lastly we would like to thank all the staff members, colleagues, non teaching staff and all our friends for their help and support from time to time.

Miss Lokhande jyoti
Mr Kamble Tushar
Mr Devare Rahule
Mr Jawale Kiran

REFERENCES

[1] L. OGorman, The document spectrum for page layout analysis, IEEE Trans. Patt. Anal. Mach. Intell., vol. 15, no. 11, pp. 11621173, Nov. 1993.

[2] B. Gatos, N. Stamatopoulos, and G. Louloudis, ICDAR 2009 handwriting segmentation contest, in Proc. Int. Conf. Document Analysis and Recognition (ICDAR), 2009, pp. 13931397.

[3] N. Stamatopoulos, B. Gatos, G. Louloudis, U. Pal, and A. Alaei, ICDAR 2013 handwriting segmentation contest, in proc. Int. Conf. Document Analysis and Recognition (ICDAR), 2013, pp. 14021406.

[4] R. Bozinovic and S. Srihari, Off-line cursive script word recognition, IEEE Trans. Patt. Anal. Mach. Intell., vol. 11, no. 1, pp. 6883, Jan. 1989.

[5] U. Mahadevan and R. Nagabushnam, Gap metrics for word separation in handwritten lines, in proc. Int. Conf. Document Analysis and Recognition (ICDAR), 1995, pp. 124127.

[6] G. Seni and E. Cohen, External word segmentation of off-line handwritten text lines, Patt. Recognit., vol. 27, no. 1, pp. 4152, Jan. 1994.

[7] V. Papavassiliou, T. Stafylakis, V. Katsouros, and G. Carayannis, Handwritten document image segmentation into text lines and words, Patt. Recognit., vol. 43, no. 1, pp. 369377, Jan. 2010.