IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

A SYSTEMATIC APPROACH TO DEVNAGARI CHARACTER RECOGNITION METHOD

Mr.Kuldeep P. Pawar*, Mr.Digvijay J. Pawar, Mr.Yashawant S. Jagadale

* Asst. Professor, Department of Electronics and Telecommunication Engineering Bharati Vidyapeeth's College of Engineering, Kolhapur (M.S.).

Asst. Professor, Department of Electronics and Telecommunication Engineering Bharati Vidyapeeth's College of Engineering, Kolhapur (M.S.).

Asst. Professor, Department of Electronics and Telecommunication Engineering Bharati Vidyapeeth's College of Engineering, Kolhapur (M.S.).

ABSTRACT

This paper provides a systematic study of segmentation methods for recognition of Devnagari characters. In this digital world, most of the documents are required in digitized manner. With the rapid growth of camera-based applications readily available on smart phones and portable devices, understanding the pictures taken by these devices semantically has gained increasing attention from the computer vision community in recent years. Among all the information contained in the image, text, which carries semantic information, could provide valuable cues about the content of the image and thus is very important for human as well as computer to understand the scenes. Text recognition is very important for humans to understand the scenes. In fact, text recognition is indispensable for a lot of applications such as automatic sign reading, language translation, navigation, and so on.

KEYWORDS:.

INTRODUCTION

In recent years research towards Indian handwritten character recognition is getting increasing attention. Devanagari script is very popular language and has very typical characteristics different from other scripts, particularly English. Combination of basic Devanagari consonants and vowels in multivariegated ways can yield as many as 100s of characters.

Character Recognition is the electronic conversion of scanned images of printed or handwritten text into machine readable text. Character recognition system is the base for many different types of applications in various fields, many of which we use in our daily lives. Devnagari character recognition is the challenging problem in Pattern Recognition and Neural Networks is one of the most commonly used techniques for character recognition and classification due to their learning and generalization abilities. A good text recognizer has many commercial and practical applications such as processing cheques in banks, documentation of library materials, extracting data from paper documents, searching data in scanned book, automation of any organization like post office, which involve lot of manual task of interpreting text.

Properties of Devnagari Script:
Devanagari script has features different from other languages. Devanagari character set has 13 vowels, 36 consonants and 10 numerals with optional modifier symbols. Characters are organized into three zones as upper, middle and lower zone. Core characters are positioned in middle zone, while optional modifiers in upper and lower zones. Two characters may be connected to each other. In Devanagari script, the concept of uppercase and lowercase characters is absent. Fig.1 represents Devanagari character set. It represents Devanagari character modifier set.Modifiers are optional symbols arranged in upper and lower zones.

Vowels		Consonants				
अ	a	क् k	ख् kh	ग् g	घ् gh	ड़ः n
आ (T)	A	च् c	छ् ch	ज् j	झ् jh	ञ् n
इ (f)	i	ट् T	ठ् Th	ड् D	ढ् Dh	ण् N
ई (T)	I	त् t	थ् th	द् d	ध् dh	न् n
उ (ु)	u	प् p	फ् ph	ब् b	भ् bh	म् m
ऊ (ू)	U	य् y	र् r	ल् l	व् v/w	
ऋ (ृ)	R	श् S	ष् sh	स् s	ह् h	ज्ञ् jn
ॠ (ॄ)	RR					
ऌ (ॢ)	lr.					
ए (े)	e					
ऐ (ै)	ai					
ओ (ो)	o					
औ (ौ)	au					
अं (ं)	m.					
अः (:)	:					

Fig 1: Devnagari Character Set

Some reasons that cause recognition of Devnagari characters difficult are as:
In Devnagari Script individual characters are connected by header line which makes segmentation of individual character is quite difficult. Characters may be connected to form conjuncts for which separation is complex. Presence of modifiers makes segmentation difficult.

Some Devanagari characters are similar in shape.

Also the unconstrained lighting conditions, various fonts, deformations, occlusions, sometimes low resolution, and complex background of text in natural scene images, the performance of scene text recognition is still unsatisfactory. Due to the high degree of intra-class variation of scene characters as well as the complexity of background, recognizing these text images is quite challenging even for state-of-the-art OCR methods.
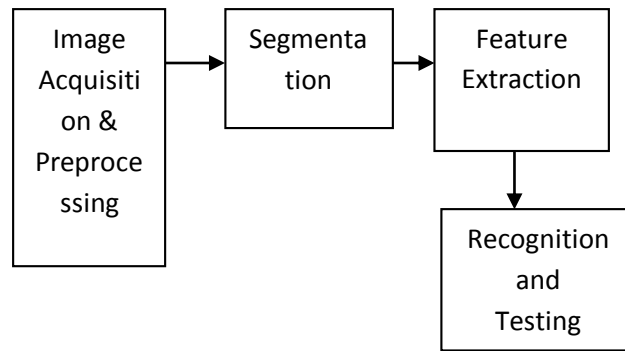
In fact, characters are designed by humans and each category of characters has unique structure representing itself. Therefore, no matter how the background changes or the character degrades, as long as the structure remains unchangeable, we could recognize them by detecting the unique structure from cluttered background. In other words, humans naturally make use of character-specific structure information when recognizing characters from scene images. Thus, a good scene character-recognition method should make use of both the local appearance and global structure information.

The objective of feature extraction is to capture the essential characteristics of the symbols and this is one of the most difficult problems of pattern recognition. In this approach, statistical distribution of points is analyzed and orthogonal properties are extracted. For each symbol a feature vector is calculated and stored in database, and recognition is performed by finding distance of feature vector of input image with those stored in the database and giving the symbol with minimum deviation. This is very sensitive to noise and edge thickness, but performs well on handwritten character set.

SYSTEM OVERVIEW
The proposed work aims for development of Devnagari text recognition system which is capable of recognizing the Devnagari characters and convert into words or sentences.

This approach is based on mainly 5 steps: Image Acquisition, Preprocessing, segmentation, Feature Extraction, Testing.

```
┌──────────┐      ┌──────────┐      ┌──────────┐
│  Image   │      │Segmenta  │      │ Feature  │
│Acquisiti │─────▶│  tion    │─────▶│Extraction│
│  on &    │      │          │      │          │
│Preproce  │      └──────────┘      └────┬─────┘
│  ssing   │                             │
│          │                             ▼
│          │                        ┌──────────┐
│          │                        │Recognition│
│          │                        │   and    │
│          │                        │ Testing  │
└──────────┘                        └──────────┘
```

Fig 2: System Overview

Preprocessing:
Image pre-processing can significantly increase the reliability of an optical inspection. Several filter operations which intensify or reduce certain image details enable an easier or faster evaluation. Users are able to optimize a camera image with just a few clicks.

Examples:
- Normalization
- Edge filters
- Soft focus, selective focus
- User-specific filter
- Static/dynamic binarisation
- Image plane separation
- Binning

Processing of Image Includes
- Includes several functions for image processing
- Contrast increase by static or dynamic binarisation, look-up tables or image plane separation
- Resolution reduction via binning
- Image rotation
- Conversion of color images to gray value images

Segmentation:
Once Image preprocessing is done it is necessary to segment document into lines, line into words and word into characters. When individual character has been separated from document we can extract features from it for recognition.

Feature Extraction:
Feature extraction is the process to retrieve the most important data from the raw data. The most important data means that's on the basis of that's the characters can be represented accurately. To store the different features of a character, the different classes are made. The feature of a character is stored to that's particular class to which it belongs [9]. So each class is different from another class. There are different methods which can be used for features extraction.

These methods are classified mainly in the three groups.
• Statistical Features
• Geometrical and Topological Features
• Global Transformation and Series Expansion.

Recognition:
Humans recognize a multitude of objects in images with little effort, despite the fact that the image of the objects may vary somewhat in different viewpoints, in many different sizes and scales or even when they are translated or rotated. Objects can even be recognized when they are partially obstructed from view.

Optical Character Recognition widely use the pattern Recognition techniques. There are also the different techniques which are investigated by the researchers. A survey report on feature extraction and classification methods for Devnagari character recognition can be found in [8].

There are the different classification techniques for Optical character recognition.
• Neural Networks.
• Statistical Techniques.
• Template Matching.
• Support Vector Machine (SVM) algorithms.
• Combination classifier.

The above techniques are not purely independent. The one technique can be considered as the subset of another.

CONCLUSION
The different methods have been proposed by the Researchers for treating the problem of Devnagari character. But a lot of research is also required to handle the challenges in Devnagari Character Recognition. It is hoped that this detailed discussion will be beneficial insight into various concepts involved, and boost further advances in the area. The accurate recognition is directly depending on the nature of the material to be read and by its quality. Current research is not directly concern to the characters, but also words and phrases, and even the complete documents. For the character recognition, HMM, neural networks and their combinations are used as the powerful tools. For the high reliability in character recognition, segmentation and classification have to be treated in an integrated manner to obtain more accuracy in complex cases. This paper has focused on an appreciation of principles and methods. The compare of effectiveness of various algorithms has not been attempted in the present work. Unfortunately there is little experimental as well as standard handwritten character database available publicly for benchmarking the accuracy of various advanced techniques proposed in Devnagari character recognition.

REFERENCES
[1] G. Sinha, Mrs. R. Rani, Prof. R. Dhir ,Recognition Using Zonal Based Feature Extraction Method and SVM Classifier ,International Journal of Advanced Research in Computer Science and Software Engineering , ISSN: 2277 128X, Volume 2,Issue 6, June 2012.

[2] M. Patil, V. Narawade, Recognition of Handwritten Devanagari Characters through Segmentation and Artificial neural networks , International Journal of Engineering Research and Technology (IJERT) , ISSN:2278-0181,Vol. 1 Issue 6, August 2012.

[3] K. Y. Rajput, S. Mishra, Recognition and Editing of Devanagari Handwriting Using Neural Network, Proceedings of SPIT-IEEE Colloquium and International Conference, Vol. 1.

[4] S. Arora, D. Bhattacharjee, M. Nasipuri, L. Malik and B. Portier, A Two Stage Classification Approach for Handwritten Devanagari Characters.

[5] V. Agnihotri, Offline Handwritten Devanagari Script Recognition, IJITCS, 2012, 8, 37-42.

[6] D. Singh, S. Singh and Dr. M. Dutta, Handwritten Character Recognition Using Twelve Directional Feature Input and Neural Network, International Journal of Computer Applications, 0975-8887, Vol. 1.

[7] N. Sharma, U. Pal, F. Kimura and S. Pal, Recognition of Off-Line Handwritten Devanagari Characters Using Quadratic Classifier.

[8] V. Bansal, R. Sinha, Segmentation of touching and fused Devanagari characters , Pattern Recognition 35,875-893, 2002.

[9] I. S. Oh, J. S. Lee, C. Y. Suen, "Analysis of class separation and Combination of Class-Dependent Features for Handwriting Recognition", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.21, no.10, pp.1089-1094, 1999.

[10] S.V. Rajashekararadhya, P. Vanaja Ranjan, "A Novel Zone Based Feature Extraction Algorithm for Handwritten Numeral Recognition of Four Indian Scripts", Digital Technology, Journal, Vol. 2, pp. 41-51, 2009.

[11] D. Trier, A. K. Jain, T. Text, "Feature Extraction Method for Character Recognition - A Survey", Pattern recognition, vol.29, no.4, pp.641-662, 1996.

[12] G. G. Rajput, S. M. Mali," Fourier Descriptor based Isolated Marathi Handwritten Numeral Recognition", Int'l Journal of Computer Applications, Volume 3 – No.4, June 2010.

[13] Sandhya Arora et al., "Performance Comparison of SVM and ANN for Handwritten Devnagari Character Recognition", IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 3, May 2010.

[14] K. Y. Rajput and Sangeeta Mishra, "Recognition and Editing of Devnagari Handwriting Using Neural Network", SPITIEEE Colloquium and Intl. Conference, Mumbai, India.