Effective Thinning Algorithm for Recognition of Hand Written Devnagri Compound Characters Using Neural Network

Prashant Yawalkar [1], Dr. M. U. Kharat [2]

[1]Research Scholar, Department of Computer Engineering, MET's Institute of Engineering, BKC Nasik, India.
[2]Professor and Head, Department of Computer Engineering, MET's Institute of Engineering, BKC Nasik, India.

Abstract

Handwriting always works as a primary tool in the world of communication. Various techniques have been and are developed in order to document the handwritten text. There are certain exceptional techniques that describe the environment of handwritten scripts and further convert it into electronic data by implementing various algorithms. Devnagri is widely used for most popular and commonly used languages like Marathi and Hindi We have proposed a neural network based hand written Devnagri compound character recognition system using a novel rule-based thinning algorithm that improves the overall performance of the system. The distinctive feature of our thinning algorithm is that it is a rotation invariant method which thins characters to their central lines helping in preserving shape of the character. The system applies different rules concurrently to each pixel in the image which results in to symmetrical thinning and improvement in the overall speed of the system. Results obtained for the system prove that the system is efficient enough to preserve the topology of the compound characters written in Devnagri that further helps in improving the accuracy of the recognition system.

Keywords: Character recognition, Neural Network, Normalization, Rotation invariant, Compound character, Thinning, Feed forward, Boundary detection, Image acquisition, Segmentation, Feature extraction.

INTRODUCTION

If handwritten characters are precise and uncontaminated, they can be recognized by humans accurately. But if the same task is given to machines, then it becomes difficult for them to do it [1]. Different languages use different scripts to write, one of them is Devnagri which is widely used for most commonly used languages like Marathi and Hindi. Presence of complicated curves and various shapes increases the complexity of recognition for these languages. All these features make Optical Character recognition (OCR) for Devnagri script specifically hard [1].

Optical character recognition [5] can be termed as the motorized or else electronic conversion of scan imagery of handwritten, typewritten or printed wording interested in system-encoded content. This process can be used for converting books and documents into electronics form, managing a record-keeping system in an office, or publishing the text on a website. Using OCR we can also modify the text, search for a phrase, store it efficiently, display a replica free of scanning artifacts, as well as use processes like machine translation, text-to-speech and text mining to it.

DEVANAGARI SCRIPT

Devanagari script is unique in a number of ways as compared to other scripts. It has two-dimensional compositions of symbols: foundation characters in the central strip as well as optional modifiers. Two characters possibly will be in shadow of every one. Whereas line segments [31] are the main features used for English, the majority of the characters inside Devnagri script are formed through curves, holes as well as strokes. In Devnagri language script, the formation of uppercase and lower-case characters is not available, but it has more number of symbols than that of English. Marathi is an Indo-Aryan language mostly spoken by millions of peoples in the western and central India [8]. Marathi is one of the Prakrit languages developed from Sanskrit and is also the official language in Maharashtra state. Typically, the handwriting manner changes from person to person as well as signature also vary from person to person [30].Having a large character set with cursive also [16], curves as well as lines are in the particular shape formation, which may be over lapping in a word. As per individual writing style a touching characters can touch each other at different position. The set of vowels and consonants in Devanagari scripts are as follows [9], [10].

Vowels	अ आ इ ई उ ऊ ऋ ए ऐ ओ औ अं अ:
Consonents	क ख ग घ ङ प च छ ज झ ञ स ट ठ ड ढ ण ह त थ द ध न क्ष प फ ब भ म त्र य र ल व श ज्ञ

Figure 1. Vowels and Consonants

FEATURES OF THE DEVANAGARI CHARACTERS:

1. Characters having End bar: Characters ending with a | are end bar characters. Eg.
   अ ख ग च ज झ त थ ध न प व भ म य ल व श स

2. Characters having No bar: Characters that do not have | are non bar characters. Eg.
   इ उ ऋ ए ट ठ ड द ड ढ ।ह