

A Review of Odia Character Recognition and Classification applied in Structural Strokes

Mr. Sandip Kumar bala, Mr. Prasanna Kumar Mishra

Department of Master in Computer Application, College Of Engineering Bhubaneswar, Odisha, INDIA.

ABSTRACT: The process of identifying a printed or handwritten document into a format that is computer readable is called character recognition. Pre-processing, segmentation, feature extraction, classification, and finally real recognition are the steps that make up a character recognition system. Most people would agree that choosing the right set of features to represent input samples is one of the key aspects affecting performance in character recognition. A review of structural strokes in character recognition is given by this paper. The core of a character recognition system is feature extraction. The various structural stroke types that are present in the character, such as ellipses, circles, top and bottom parts of circles, and horizontal and vertical strokes, form the basis for the character's features. The character's structure is made up of these symbol combinations. Not every symbol exists in a character; it all relies on how the writing style shapes the symbol. There have been several works done on various characters, such as Hindi, Urdu, English, etc., but not many on the odia character. Numerous approaches have been put out thus far for the recognition of the odia character. This report attempts to examine the research that has been conducted over the past few years on this ancient character by different researchers.

KEYWORDS: Character recognition, odia character, feature extraction, structural strokes.

1. INTRODUCTION: The process of converting a printed or handwritten document into a machine-readable format is called character recognition. Three essential processes make up character recognition: feature extraction, classification, and pre-processing. Based on access mechanisms, character recognition is divided into two classes: handwritten and machine printed. Characters produced by machines are consistent, recognizable, and Computers can recognize handwritten characters and make them comprehensible. Although there are many distinct kinds of languages spoken in India, the Indian Constitution recognizes 24 official languages. English and the state official language are the two primary categories of official languages. The Brahmin script is where this odia character got its start. The odia character is composed of multiple connected symbols, such as circular, vertical, and horizontal. Character recognition is the process of transforming a handwritten or printed document into a format that can be read by a machine. Character recognition involves three crucial steps: pre-processing, classification, and feature extraction. Character recognition is separated into two classes: machine printed and handwritten, based on access mechanisms. Computers can recognize handwritten characters and interpret them, and they can produce consistent, recognizable characters. The Indian Constitution recognizes 24 official languages, despite the fact that there are many different languages spoken in the country. The two main categories of official languages are English and the official language of the state. This particular odia character originated in the Brahmin script. The odia character is made up of several interconnected symbols, including horizontal, vertical, and circular ones.

OVERVIEW OF OCR PROCESS:

The process of converting a handwritten or printed document into a machine-readable format is known as Odia character recognition, or OCR. The writing style (stroke), which is the movement of the writing instrument on the writing surface, determines the character. In Ovi script, characters are arranged from bottom to top and have a circular shape, with many having a similar appearance. The following crucial procedures are used by the odia character recognition system.

PRE-PROCESSING: Pre-processing is one of the first steps in character recognition. In order to obtain accurate results, feature extraction is performed prior to using the row data. The following pre-processing steps involve working with scanned input images. operations like skew correction, inversion, slant removal, thresholding, noise reduction, binarization, and strokes with normalization .

SEGMENTATION: Several approaches have also been used in segmentation for the script, taking into account word choice, character development, and line breaks. The main difficulty in this process is separating the lines for script identification, which is done to make it easier to extract each word segment. Hough Transform, horizontal projection, and smearing etcmentation are based on vertical projection and the connection component in text line segmentation techniques .

FEATURE EXTRACTION: Every character in feature extraction has a unique characteristic that makes it stand out from the others. Feature extraction is the process of identifying a set of parameters that define a character in a unique way. By extracting as many features as possible for each character from the sample image, this method increases the likelihood that the character will be recognized. Every character is represented by a feature vector, which serves as the character's identity. Statistical feature extraction and structural feature extraction are the two main feature extraction techniques used in this method .

CLASSIFICATON:

It feeds the classifier with the feature that was extracted in the earlier step. To assign a class to an input, the classifier compares it with the stored feature. There are several methods for classifying data: kernel, artificial neural network (ANN), statistics, and multiple classifier combinations. The recognition of characters or words is the classification stage. The classification stage would use the data to identify the feature class based on the properties in the features after features that represent the raw input data are extracted. The classification approach has a wide range of applicable techniques. Template matching, statistical methods, syntactic analysis, and neural networks can all be linked to the classification process .Classification of the character: There are many different ways by which the image can be classified. Following are some of the classifiers which are usually used for recognition of features. Apart from this there are many other classifiers available which can be used for specific classification tasks. Each has their own merits and demerits. Their accuracy also varies from each other by a large factor.

SUPPORT VECTOR MACHNE:

The identification of the character There are two classes for the character: training and testing states. In the training state, every character has a single target value along with a number of corresponding features. the target character by contrasting it with the qualities of the character. It is an accurate classifier, among the best.

HIDDEN MARKOV MODEL

We are using some unobserved (hidden) states to train the system in this statistical model.It is made up of a limited number of states, each of which has a probability distribution attached to it.A result or observation can be produced in a specific state based on the related probability distribution.Because the matrix matrix is constant and does not change over time, the state transition matrix is independent of time.This model is known as the Hidden mark over model because it only allows us to observe the outcomes in which the states are hidden from the view of an outside observer .

ARTIFICIAL NEURAL NETWORK

It is a computational or mathematical model that mimics the composition and operation of a biological neural network. It is an adaptive system, and during the learning phase, the structure of the network is dependent upon internal or external information. One input layer, multiple output layers, and a few intermediate layers make up an ANN system. All other intermediate layers are hidden, with the exception of the final output layer. The input layer's output is supplied to the hidden layers, which in turn supply the output layer with their own output. It is possible to train the hidden layers to produce an appropriate output image.

2. ODIA LITERATURE:

Odia is one of the most widely used and ancient scripts. Kalinga script, descended from Brahmin script, gave rise to Odia script. Odia script is circular in nature and is used to write the Odia language. Both simple and complex characters can be found in the current Odia script. In Odia, there are twelve vowels, thirty-seven simple consonants, ten numerical digits, and almost two hundred composite characters (juktas). The writing style for this Odia character is from left to right. Odia's character shapes are primarily circular. Occasionally, a consonant character will combine with another consonant to form a new character called a matra. In Odia script, a vowel is followed by a matras, which can be positioned at the bottom, right, or left of the vowel.

3. ROBUST STRUCTURAL SOLUTION:

Sukadev Meher and Debasis Basu : The Odia symbol was composed of multiple interconnected symbols. The approach suggested in the text is divided into line development for character recognition. The text is divided into lines, which are further divided into individual words, and words are further divided into symbols. Symbols are recognized as a basic unit of character recognition. Character recognition was made easy by the character's structural technique.

4. SUPPORT VECTOR MACHINE:

Nandni Das Behera, Himadri, and Sanghamitra Mohanti: This method differentiates characters line by line. The upper, lower, and horizontal projection profiles for the pixel intensity in each zone are connected to the matras along with the line height. The single character recognition system lowers the accuracy rate. The character is divided into two classes: training and testing states. Every character stroke in the training state has a single target value and several associated features. The target character by drawing comparisons between it and the attributes of the character. It is one of the best classifiers in terms of accuracy.

5. ENVISAGING METHOD:

The recognition of handwritten Odia characters focuses on developing strategies that can effectively extract the feature vector from each individual character. The primary task for this topic is feature extraction and recognition, with an accuracy of 70% for the final result. The characters are examined in the database's feature vector of the picture matrix form.

6. COMBINING TRAIL METHOD:

The technique for recognizing character strokes uses this method. The characters are used in a quantity of strokes, with each stroke being divided and examined separately. This technique detects the strokes of characters with an accuracy of 80%.

7. STRUCTURAL AND STATISTICAL METHOD:

The number of horizontal or vertical lines, the number of end points, the number of cross points, the horizontal curve at the top or bottom, etc. are examples of structural features in the structural method. Based on the topological and geometrical characteristics of a character, this type structure determines the structural characteristics of a character. The statistical method determines a character's statistical feature. The statistical distribution of the pixels is the source of the statistical features. When compared to structural features, these features are more easily detected. For zoning, projection histograms, crossing, and distance, among other purposes, feature extraction is employed. It is simple to extract the character

feature using this method.

8. FEATURE EXTRACTON TECHNIQUE:

This method is used to monitor the character structural strokes in character. The structure is the intensity with noise and the whole in the mask are used in feature extraction of strokes technique.

9. BI-QUADRATIC INTERPOLATION METHOH:

This technique is used for character recognition using the curve like stroke detection method. That's curve is used in reference of recognition purpose and Quantized in three level according to curve. The gradient is quantized in to 32 level and pay/16 intervals and the strength of the gradient is accumulated in each of the direction and each of the 3 curvature levels of every blocks. This method is used in circular type characters stroke detection method.

10. FEATURE EXTRACTION TECHNIQUE:

The primary idea is to extract the character counter's geometric features and create the character skeleton by basing it on the most common line types. A neural network trained on the feature vectors was used to test the feature extraction technique explain.

11. CONCLUSION:-

An extensive and current review of odia character recognition is presented in this paper. A survey has been conducted on the research work done in the last decade in the areas of stroke detection and odia character recognition. The various methods used at each stage of ODA character recognition are also described. Every one of these approaches has pros and cons of its own. Character recognition is still a difficult scientific problem because different types of strokes and different types of structures can produce variations in a character that looks the same. In the past few decades, extensive research studies have been conducted to identify characters and structural strokes in a variety of Indian and foreign languages; however, only a small amount of work has been reported on the recognition of odia characters.

12. REFERENCES:-

- [1] Debasis Basa¹, Sukadev Meher², "Handwritten Odia Character Recognition" Department of Electronics and Communication, National Institute of Technology, Rourkela, India.
- [2] Sanghamitra Mohanty, Himadri Nandini Das Bebartta. "A Novel Approach for Bilingual (English - Oriya) Script Identification and Recognition in a Printed Document", International journal of Image processing (IJIP), Volume no-(4), Issue no-(2).
- [3] Anuraag Hota. Souramya Pradhan, Hand written odia character recognition", Dept. Of electronic and communication engineering. NIT Rourkela-2012.
- [4] Yan Xu, Xiangnian Huang, Huan Chen, Huizhu Jiang, "A New Method for Chinese Character Strokes Recognition" Xihua University, Chengdu, China Email: 501625379@qq.com Received July 6, 2012; revised August 4, 2012; accepted August 16, 2012 .
- [5] Purna Vithlani, C.K.Kumbharana, "Structural and statistical feature extraction methods for character and digit recognition", International Journal of Computer Applications (0975-8887), volume 120-no. 24-Jun 2015.
- [6] "Odia Offline Character Recognition", Priyaranjan Behera. 108CS021, B.Tech.(CSE), 2008-12, Department of Computer Science and Engineering. National Institute of Technology, NIT Rourkela.
- [7] Hung-Pin Chiu, Din-Chang Tseng*. "A novel stroke-based feature extraction for handwritten Chinese character recognition", Pattern Recognition 32 (1999) 1947-1959.