Identification and Conversion of Handwritten Malayalam Scripts Using convolutional Neural Networks

[1]Joslin Johnson, [2]Ashly Raphel, [3]Catherine Davis, [4]Mrs Soumya Varma

[1,2,3]Research Scholars, [4]Assistant Professor
Department of Computer Science & Technology
Sahrdaya College of Engineering & Technology, Kerala

***ABSTRACT*: Handwritten character recognition is still a research challenge in OCR discipline, especially for Indian scripts. The recognition of handwriting however, is still considered as an open research problem due to its substantial variation in appearance. The goal of OCR is to classify optical patterns in an image to the corresponding characters. There are different OCR techniques like Online Handwritten Malayalam Character Recognition using LIBSVM in Matlab. Here real time (x,y) coordinates per stroke are acquired and preprocessed. Directional and Curvature features are extracted and trained in LIBSVM, a tool for SVM Classifers. Testing alphabet is given online to the trained SVM network and the recognized label is displayed in Notepad. Another method proposed is a two-stage approach. The first stage is a group classifier, where a group consists of similar characters and those that misclassify among themselves. In the second stage, a character assigned to a group in the first stage is classified to a particular character class. Another method proposed is Off-line Handwritten Character Recognition using Hidden Markov Model. Training and recognition are performed using Hidden Markov Model Toolkit. Recognition process involves several steps including image acquisition, dataset preparation, pre-processing, feature extraction, training and recognition. An average accuracy of about 81.38% has been obtained. Thus the proposed system is done in python using convolution neural network (CNN). It takes less processing time and has an accuracy of 95%.**

KEYWORDS: Optical Character Recognition, Support Vector Machine, Convolutional Neural Network, Machine Learning

INTRODUCTION

In the running world, there is growing demand for the software systems to recognize characters in computer system when information is scanned through paper documents as we know that we have number of newspapers and books which are in printed format related to different subjects. These days there is a huge demand in "storing the information available in these paper documents in to a computer storage disk and then later reusing this information by searching process". One simple way to store information in these paper documents in to computer system is to first scan the documents and then store them as images. But to reuse this information it is very difficult to read the individual contents and searching the contents form these documents line-by-line and word-by-word. The reason for this difficulty is the font characteristics of the characters in paper documents are different to font of the characters in computer system. As a result, computer is unable to recognize the characters while reading them. This concept of storing the contents of paper documents in computer storage place and then reading and searching the content is called document processing. Sometimes in this document processing we need to process the information that is related to languages other than the English in the world. For this document processing we need a software system called **CHARACTER RECOGNITION SYSTEM**. This process is also called document image analysis.

Thus our need is to develop character recognition software system to perform Document Image Analysis which transforms documents in paper format to electronic format. For this process there are various techniques in the world. Among all those techniques we have chosen Optical Character Recognition as main fundamental technique to recognize characters. The conversion of paper documents in to electronic format is an on-going task in many of the organizations particularly in Research and Development (R&D) area, in large business enterprises, in government institutions, so on. From our problem statement we can introduce the necessity of Optical Character Recognition in mobile electronic devices such as cell phones, digital cameras to acquire images and recognize them as a part of face recognition and validation.
To effectively use Optical Character Recognition for character recognition in-order to perform Document Image Analysis, we are using the information in Grid format. . This system is thus effective and useful in virtual digital library's design and construction.

OCR is one of the most challenging areas of image processing and pattern recognition. It plays a vital role in creating digital library expanded. It is highly essential and unavoidable while dealing with Indian languages for which there has been little digital access. Only few approaches had been devised for handwritten Malayalam documents which include wavelet Transforms, Kohonen Networks and Projection Profiles. Since little attempts have been made to develop OCR that could recognize handwritten Malayalam documents, this area needs further more developments and the researches are still going on this field. A lot of techniques of pattern recognition such as Template Matching, Neural Networks, Syntactical Analysis, Hidden Markov Models, Bayesian Theory, etc have been exhumed to develop robust OCRs for different languages. The current system has efficient and inexpensive OCR packages which are commercially available for the recognition of printed and handwritten

documents. Among those we have enough facilities for languages such as English, Chinese etc. When considering the Indian languages, many attempts are made to develop the OCR system for Devanagari, Oriya, Tamil, Telugu, and Kannada etc. While taking Malayalam into consideration an effective method of recognition is still promising. The recognition of handwritten character recognition poses a great challenge to researchers. Even now a lot of problems in this area are still to be addressed. Handwritten character recognition (HCR) system is so complex with the variety of character structure and distorted and broken characters and personal independence.

It is hard to say that handwritten recognition exits for Malayalam language. In has proposed an algorithm for the recognition of isolated handwritten Malayalam characters which used the HLH intensity patterns for the feature extraction technique. The input used in the present work is the image input given by the Light pen device. The characters are written through Light pen device and it is converted into 24 bit bmp image. The output is an editable computer file which is the equivalent character written by the user.

Malayalam is a South Indian language - which is the principal language of the State of Kerala, spoken by about 36 million people in the world. The Malayalam script is a Brahmic script used commonly to write the Malayalam language. Like many other Indic scripts, Malayalam follows a writing system that is partially alphabetic and partially syllable-based. The Malayalam script uses both old and new script for depicting characters.

CHAPTER 2

LITERATURE SURVEY

People start learning to read and write during the early stage of education. As years pass by they may have acquired good reading and writing skills. It may not be difficult for them to read any kind of either printed or handwritten characters.But Computers may find difficulty in deciphering many kinds of printed characters which is of different fonts and styles or handwritten characters. Malayalam OCR is a complex task owing to the various character scripts available and more importantly the difference in ways in which the characters are written. The dimensions are never the same and may be never mapped onto a square grid unlike English characters. This survey paper provides the details of different malayalamocr modules and their techniques for identifying and recognizing the malayalam old scripts and converting it to new Malayalam script.

OCR is one of the most challenging areas of image processing and pattern recognition.OCR plays a vital role in creating digital library expanded. It is highly essential and unavoidable while dealing with Indian languages for which there has been little digital access. Only few approaches had been devised for handwritten Malayalam documents which include wavelet Transforms, Kohonen Networks and Projection Profiles. Since little attempts have been made to develop OCR that could recognize handwritten Malayalam documents, this area needs further more developments and the researches are still going on this field. A lot of techniques of pattern recognition such as Template Matching, Neural Networks, Syntactical Analysis, Hidden Markov Models, Bayesian Theory, etc have been exhumed to develop robust OCRs for different languages. The current system has efficient and inexpensive OCR packages which are commercially available for the recognition of printed and handwritten documents. Among those we have enough facilities for languages such as English [1], Chinese [2] etc. When considering the Indian languages, many attempts are made to develop the OCR system for Devanagari, Oriya, Tamil [3], Telugu [4], and Kannada [5] etc. While taking Malayalam into consideration an effective method of recognition is still promising. The recognition of handwritten character recognition poses a great challenge to researchers. Even now a lot of problems in this area are still to be addressed. Handwritten character recognition (HCR) system is so complex with the variety of character structure and distorted and broken characters and personal independence.

It is hard to say that a complete Malayalam OCR exits which meets all conditions. Malayalam OCR lacks an efficient algorithm. Even in the field of printed charactersthere are little advancements for this language. Even though the administrative language of Kerala is Malayalam, only a few works were reported in this area. Government of Kerala has now taken initiative for the development of this language and scope of development in this area is promising. The first OCR system was developed by Centre for Development of Advanced Thiruvananthapuram, a Government of India institution. It uses Otsu's algorithm for binarization and Projection profile method used for skew detection and correction of image. The recognition phase linguistic rules are applied. An accuracy of 97% is reported in this method. Another system is reported by M Abdul Rahiman and M S Rajasree which uses wavelet based feature extraction and neural network based recognition. Bindu Philip and R D Sudhakara Samuel proposed an OCR for Malayalam using column stochastic image matrix. In Neeba N V and C V Jawahar proposed a method of recognition of Malayalam characters from books. The recognition of handwritten Malayalam character is still in the stage of infancy. Only a little research is going on in this area. Our earlier work in the field of handwritten Malayalam character recognition provided a new method for isolated characters. HLH intensity patterns were used for the recognition of characters and an accuracy of 86 percentages was achieved. Another work was reported by G Raju in which thedaubechie wavelets (db4) were used for recognition. Lajish V L, Suneesh T K K and Narayanan N K proposed a system which is based on statistical classification. Artificial Neural Networks are applied for recognition of Handwritten Malayalam characters in the work done by Lajish V The area of handwritten Malayalam character is still promising and offers a plethora of opportunities for research.

It is hard to say that handwritten recognition exits for Malayalam language. In [6] has proposed an algorithm for the recognition of isolated handwritten Malayalam characters which used the HLH intensity patterns for the feature extraction technique. The input used in the present work is the image input given by the Light pen device. The characters are written through Light pen device and it is converted into 24 bit bmp image. The output is an editable computer file which is the equivalent character written by the user.

2.1 METHODOLOGY

Malayalam is a South Indian language - which is the principal language of the State of Kerala, spoken by about 36 million people in the world. The Malayalam script is a Brahmic script used commonly to write the Malayalam language. Like many other Indic scripts, Malayalam follows a writing system that is partially alphabetic and partially syllable-based. The Malayalam script uses both old and new script for depicting characters. Noise Removal using skew detection and projection profile has acquired the result set shown in Table 1. The proposed system has the advantage of of n-gram segmentation with an efficacy of 98% which is higher than the efficiency of uni-gram approach with 78% worth. A comparative result analysis of uni-gram and n-gram segmentation could be made considering the time taken for obtaining the outputs. uni-gram segmentation of two similar word snippets in Malayalam which takes 0.0020 seconds in recognition for each letter represents n-gram segmentation. In uni-gram isolation, n=1 by default and takes 0.012 seconds for isolating a 6 which is similar to the that has already been included in trained data set.In the case of n-gram segmentation, it takes only 0.0040 seconds to identify if the trained data already constitutes, which is less than the time required by uni-gram segmentation technique for isolating the same word snippets. Here,segmenting each letter takes 0.0020 seconds such that unigram segmentation considers each letter individually whereas n-gram segmentation, takes three times less than the time required by uni-gram isolation since n-gram considers the 6 letter word as 2 isolated character with each isolated segment equal to 5 characters(or lesser if no letter follows).

Due to the complexity of the Malayalam character set, an efficient method for the recognition for handwritten characters has not been proposed till now. Based on Otsu's algorithm for binarization an OCR system was devised by Centre for Development of Advanced Computing [7] (CDAC) Thiruvananthapuram, Kerala, a Government of India Institution. In this system, projection profile method is used for skew detection and correction of image; and in the recognition phase linguistic rules are applied. An accuracy of 97% was reported in this method. Using wavelet based feature extraction and neural network based recognition, a new work was reported by M Abdul Rahiman and Rajasree. Another work was reported by G Raju, in which the daubechie wavelets (db4) were used for recognition. Another OCR system was proposed by Lajish V L, Suneesh T K and Narayanan N K which was based on statistical classification. Most recently, a method for the recognition of Isolated Handwritten Malayalam Character using HLH Intensity Patterns was devised by M Abdul Rahiman, G Manoj Kumar and M S Rajasree .

CHAPTER 3

SYSTEM ANALYSIS

The system can be analyzed mainly by two things existing system and proposed system.

3.1. EXISTING SYSTEM

Due to the complexity of the Malayalam character set, an efficient method for the recognition for handwritten characters has not been proposed till now. Based on Otsu's algorithm for binarization an OCR system was devised by Centre for Development of Advanced Computing (CDAC) Thiruvananthapuram, Kerala, a Government of India Institution. In this system, projection profile method is used for skew detection and correction of image; and in the recognition phase linguistic rules are applied. An accuracy of 97% was reported in this method. Using wavelet based feature extraction and neural network based recognition, a new work was reported by M Abdul Rahiman and Rajasree. Another work was reported by G Raju, in which the daubechie wavelets (db4) were used for recognition. Another OCR system was proposed which was based on statistical classification. Most recently, a method for the recognition of Isolated Handwritten Malayalam Character using HLH Intensity Patterns was devise. This method employs recognition of isolated handwritten characters in a noiseless environment. The basic principle is 2010 Second International Conference on Machine Learning and Computing to identify specific terminologies in each character and extend the same to a set of characters in order to achieve accurate results with very low complexity algorithms. The separation letters is shown in which uses intensity variations for segregating the line and character from the scanned image.

3.2. PROPOSED SYSTEM

In the proposed system we are attempting to convert all handwritten malayalam old scripts (hard copies) to its editable form (soft copies). This is more useful in Government offices where a number of documents have to be handled. In such offices maintaining a soft copy of the documents is more organized than keeping a hardcopy, especially for old documents. Therefore using this system we can keep soft copies of all these documents and hence the problem of damaged old documents can be avoided. In this project, we are trying to identify and convert old scripts of malayalam language to its present age new scripts using convolutional neural network. First the system is trained with different Malayalam characters written in different styles. Then image of the malayalam old scripts is given as input to the system. Then noise is removed from the scanned image. Then the segmentation process segments each character in the image. Each segmented character is the recognized and then converted to text which is the new malayalam script. This method employs recognition of isolated an combinational handwritten characters in a noiseless environment. The basic principle is to identify specific terminologies in each character and extend the same to a set of characters in order to achieve accurate results with very low complexity algorithms. This work separate the entire character set in to three different classes. Ra type characters, Pa type characters and Special symbols. This classification is based on the shape and appearance of the character. This shape feature is extracted to recognize the letter.

Once the segregation is accomplished, the feature extraction process is initialized. The length and breadth of each character can be calculated by manipulating the HLH intensity values of the segregated image, which in turn, is stored in a dynamic window matrix Inferences are arrived at on the basis of the sequence pattern procured on horizontally processing the dynamic matrix. Furthermore, the pattern with highest probability is identified. The matrix is then processed for vertical as well as straight line patterns. Consider the intensity matrix of the extracted Malayalam character, pronounced as Ra, illustrated as in figure 2.The same character maybe depicted in different ways and with different thickness. Irrespective of this the pattern HLH is identified as the pattern with highest probability. Consider the character depicted as in figure 3. The intensity pattern HLH can be observed in the letter and hence infer that two vertical pillars exist on processing the image horizontally. This work separate the entire character set in to three different classes. Ra type characters, Pa type characters and Special symbols. This classification is based on the shape and appearance of the character. This shape feature is extracted to recognize the letter.

CHAPTER 4

DESIGN AND DEVELOPMENT

4.1. MODULE DESIGN

Optical Character Recognition, or OCR, is a technology that enables you to convert different types of documents, such as scanned paper documents, PDF files or images captured by a digital camera into editable and searchable data.

First, the program analyzes the structure of document image. It divides the page into elements such as blocks of texts, tables, images, etc. The lines are divided into words and then - into characters. Once the characters have been singled out, the program compares them with a set of pattern images. It advances numerous hypotheses about what this character is. Basing on these hypotheses the program analyzes different variants of breaking of lines into words and words into characters. After processing huge number of such probabilistic hypotheses, the program finally takes the decision, presenting you the recognized text.

In the proposed system we are attempting to convert all handwritten Malayalam old scripts (hard copies) to its editable form (soft copies). In such offices maintaining a soft copy of the documents is more organized than keeping a hardcopy, especially for old documents. Therefore using this system we can keep soft copies of all these documents and hence the problem of damaged old documents can be avoided. In this project, we are trying to identify and convert old scripts of malayalam language to its present age new scripts using OCR module. It involves the following steps. First the image of the malayalam old scripts is given as input to the system. Then noise is removed from the scanned image. Then the image is converted to text which is the new malayalam script.

Block diagram of the whole process

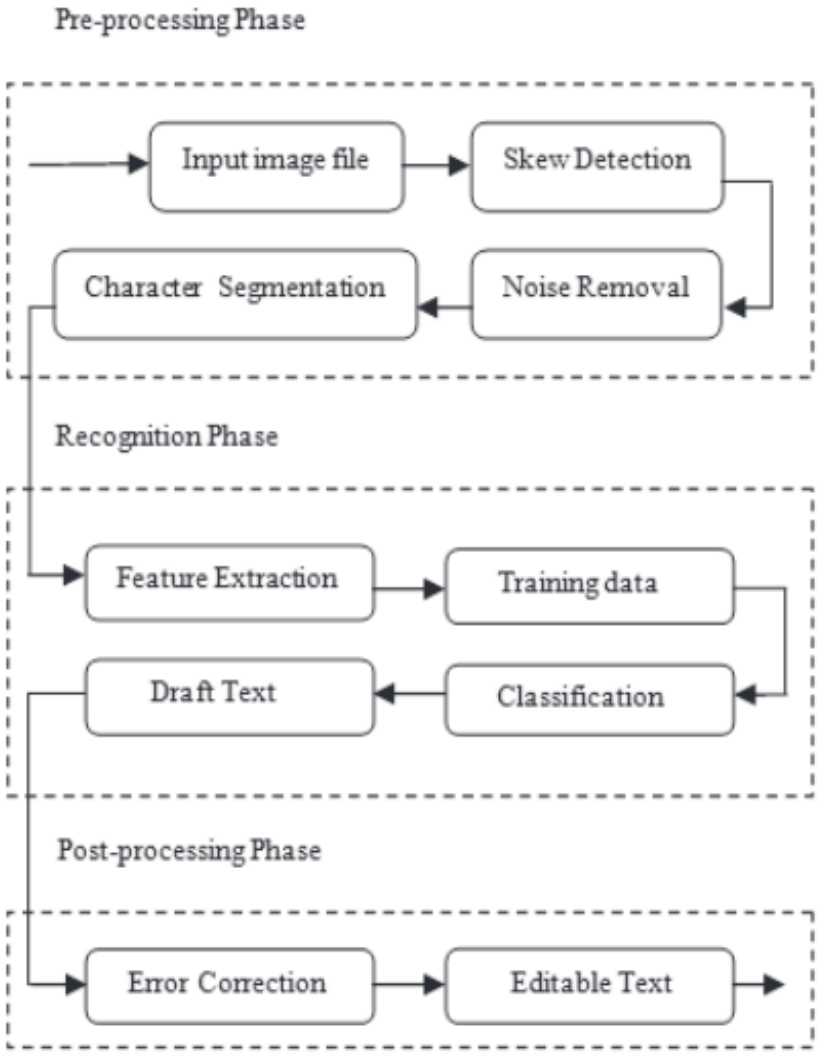

Fig 4.1: Block diagram for OCR

4.2. MODULE DESCRIPTION

The character identification is done through 3 phases:

- Pre-processing
- Skeletonization
- Recognition

4.2.1 Preprocessing Technique

Scanning of the text which converts the paper document into an image is the starting stage. The document is scanned by any standard scanners with a minimum of resolution 200 dpi. Lower resolution results in poor performance of the system and misidentifications. A scanned image will be anyone ofthe formatjpeg, bmp or tiff. This image is processed in many stages.

This technique involves the steps done before the actual identification is done. In this phase, colour of the background of the text and the written characters is checked. If they are of different colours, methods are applied to produce uniformly coloured characters. Background noises up to certain intensities can be identified and removed. In order to identify the characters, the characters are segmented first to produce individual units of characters. The scanned text is first subjected to line separation process where the written document is separated into line of characters. After line separation, each character in a line is subjected to the character separation process. Here the characters in the line are separated into individual units which simplify the processes in the following phases.

4.2.2 Skeletonization

Skeletonization algorithms are the need to compute a reduced amount of data or to simplify the shape of an object in order to find features for recognition algorithms and classifications. It is the transformation of a component of a digital image into a subset of the original component. There are different categories of Skeletonization methods: one category is based on distance transforms, and a specified subset of the transformed image is a distance skeleton. The original component can be reconstructed from the distance skeleton. Another category is defined by thinning approaches. The result of Skeletonization using thinning algorithms should be a connected set of digital curves or arcs. The segmented characters are subjected to the thinning algorithm. Thinning is the process of peeling off a pattern as many pixels as possible without affecting the general shape of the pattern. The skeleton obtained must be as thin as possible, connected and centered. Individual pixels are either removed in a sequential order or in parallel. Normally, it is implemented by an iterative process of transforming specified contour points into background points.

4.2.3 *Recognition*

The final phase of the identification of characters involves a series of methods. The skeletonised and segmented characters are made to undergo functions which calculate the number of horizontal and vertical lines which form the features of the characters. Using the count of horizontal and vertical lines, the characters are classified into different groups. For example consider the character ‘Ra’. It has two vertical lines and a single horizontal line. So it can be classified into group of characters having similar features. For recognition of certain characters, the count of horizontal and vertical lines is enough. But for other characters such as ‘La’, ‘Va’, ‘Pa’ etc the position of these lines are also important as these differentiate each other. Hence the positions of these lines are also calculated i.e. whether at the top, bottom, left or right. After calculating the count and position of horizontal and vertical lines, the characters are classified to form different group.

4.3. UML DIAGRAMS

UML stands for Unified Modeling Language which is used in object oriented software engineering. Although typically used in software engineering it is a rich language that can be used to model an application structures, behavior and even business processes. There are 14 types of UML diagrams. They can be divided into two main categories; structure diagrams and behavioral diagrams.

4.3.1. USE CASE DIAGRAM

Use case diagrams are usually referred to as behavior diagrams used to describe a set of actions (use cases) that some system or systems (subject) should or can perform in collaboration with one or more **external users** of the system (actors). Each use case should provide some observable and valuable result to the actors or other stakeholders of the system.

Use case diagram

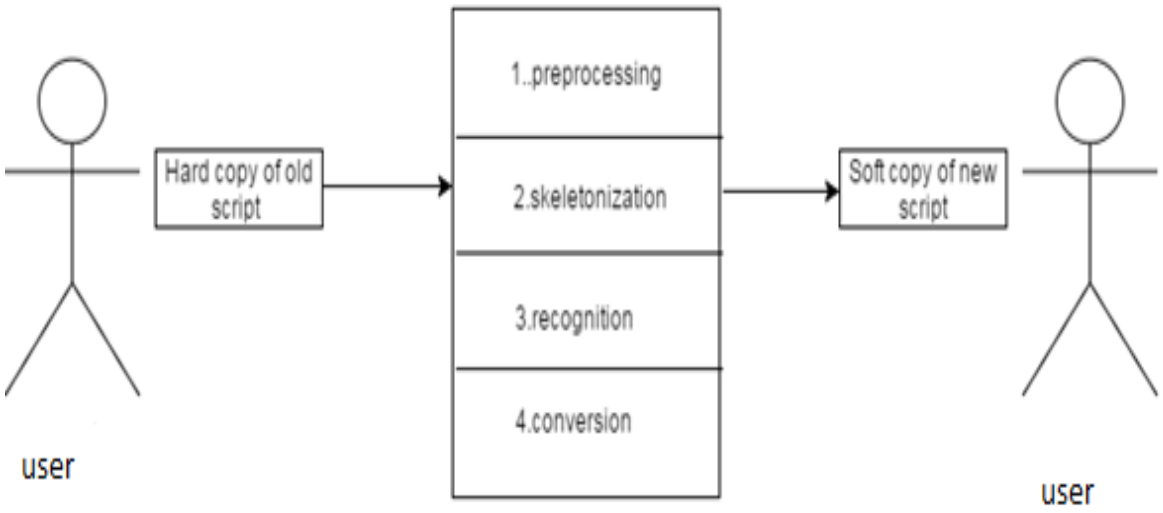

Fig 4.2: Use case diagram

4.3.2. SEQUENCE DIAGRAM

A Sequence diagram is an interaction diagram that shows how objects operate with one another and in what order. It is a construct of a message sequence chart. It shows object interactions arranged in time sequence. It depicts the objects and classes involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario. Sequence diagrams are typically associated with use case realizations in the Logical View of the system under development. Sequence diagrams are sometimes called event diagrams or event scenarios. This allows the specification of simple runtime scenarios in a graphical manner.

Sequence Diagram

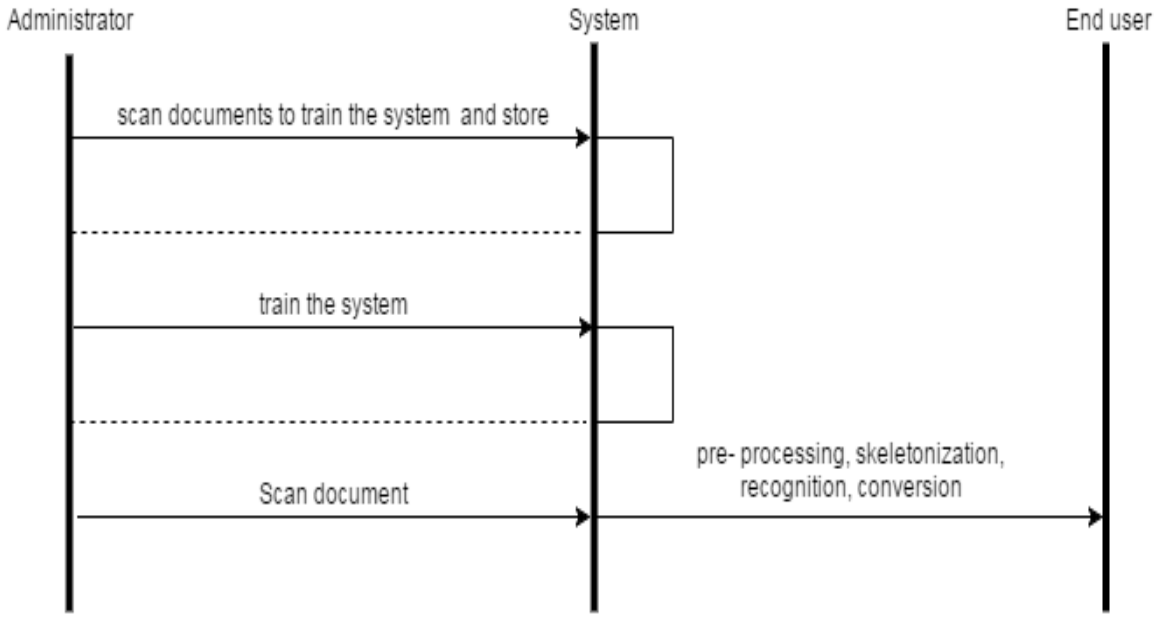

Fig 4.3: Sequence Diagram

4.3.3. FLOWCHART

A flowchart is a type of diagram that represents an algorithm, workflow or process, showing the steps as boxes of various kinds, and their order by connecting them with arrows. This diagrammatic representation illustrates a solution model to a given problem.

Flow Chart

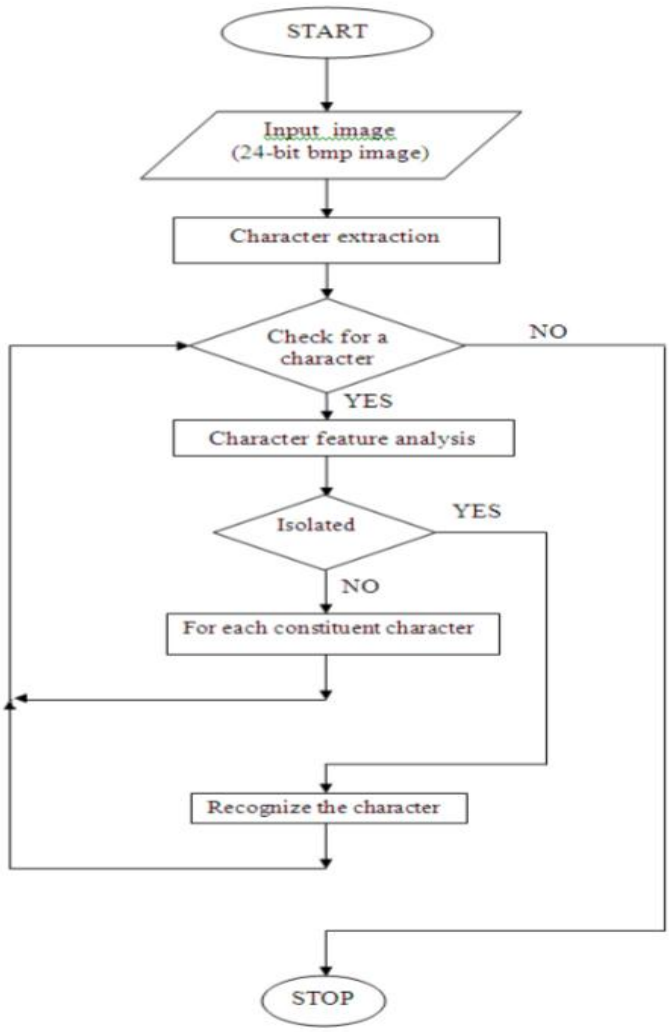

Fig. 4.4: Flow Chart

CHAPTER 5

TECHNOLOGY DESCRIPTION

5.1. HARDWARE DESCRIPTION

5.1.1. Scanner

A scanner is a device that captures images from photographic prints, posters, magazine pages, and similar sources for computer editing and display. Scanners come in hand-held, feed-in, and flatbed types and for scanning black-and-white only, or color. Very high resolution scanners are used for scanning for high-resolution printing, but lower resolution scanners are adequate for capturing images for computer display. Scanners usually come with software, such as Adobe's Photoshop product, that lets you resize and otherwise modify a captured image. Scanners usually attach to your personal computer with a Small Computer System Interface (SCSI). An application such as Photoshop uses the TWAIN program to read in the image. Some major manufacturers of scanners include: Epson, Hewlett-Packard, Microtek, and Relisys.

5.1.2. Camera

A camera is an optical instrument for recording or capturing images, which may be stored locally, transmitted to another location, or both. The images may be individual still photographs or sequences of images constituting videos or movies. The camera is a remote sensing device as it senses subjects without physical contact. The word camera comes from *camera* obscura, which means "dark chamber" and is the Latin name of the original device for projecting an image of external reality onto a flat surface. The modern photographic camera evolved from the camera obscura. The functioning of the camera is very similar to the functioning of the human eye.

5.2. SOFTWARE DESCRIPTION

5.2.1Python

Python is a widely used high-level, general-purpose, interpreted, dynamic programming language. Its design philosophy emphasizes code readability, and its syntax allows programmers to express concepts in fewer lines of code than possible in languages such as C++ or Java. The language provides constructs intended to enable writing clear programs on both a small and large scale

Python supports multiple programming paradigms, including object-oriented, imperative and functional programming or procedural styles. It features a dynamic type system and automatic memory management and has a large and comprehensive standard library.

Python interpreters are available for many operating systems, allowing Python code to run on a wide variety of systems. Using third-party tools, such as Py2exe or Pyinstaller, Python code can be packaged into stand-alone executable programs for some of the most popular operating systems, so Python-based software can be distributed to, and used on, those environments with no need to install a Python interpreter.

5.2.2 Machine Learning

Machine learning is the subfield of computer science that gives computers the ability to learn without being explicitly programmed. Evolved from the study of pattern recognition and computational learning theory in artificial intelligence,[1] machine learning explores the study and construction of algorithms that can learn from and make predictions on data[2] - such algorithms overcome following strictly static program instructions by making data driven predictions or decisions,[3]:2 through building a model from sample inputs. Machine learning is employed in a range of computing tasks where designing and programming explicit algorithms with good performance is difficult or infeasible; example applications include spam filtering, detection of network intruders or malicious insiders working towards a data breach,[4] optical character recognition (OCR),[5] search engines and computer vision.

Machine learning is closely related to (and often overlaps with) computational statistics, which also focuses in prediction-making through the use of computers. It has strong ties to mathematical optimization, which delivers methods, theory and application domains to the field. Machine learning is sometimes conflated with data mining, where the latter subfield focuses more on exploratory data analysis and is known as unsupervised learning. Machine learning can also be unsupervised and be used to learn and establish baseline behavioral profiles for various entities and then used to find meaningful anomalies.

Within the field of data analytics, machine learning is a method used to devise complex models and algorithms that lend themselves to prediction; in commercial use, this is known as predictive analytics. These analytical models allow researchers, data scientists, engineers, and analysts to "produce reliable, repeatable decisions and results" and uncover "hidden insights" through learning from historical relationships and trends in the data.

5.2.3 Convolutional Neural Network

In machine learning, a convolutional neural network (CNN, or ConvNet) is a type of feed-forward artificial neural network in which the connectivity pattern between its neurons is inspired by the organization of the animal visual cortex. Individual cortical neurons respond to stimuli in a restricted region of space known as the receptive field. The receptive fields of different neurons partially overlap such that they tile the visual field. The response of an individual neuron to stimuli within its receptive field can be approximated mathematically by a convolution operation. Convolutional networks were inspired by biological processes and are variations of multilayer perceptrons designed to use minimal amounts of preprocessing. They have wide applications in image and video recognition, recommender systems and natural language processing.

CHAPTER 6

CONCLUSION

This project helps to convert all handwritten Malayalam old scripts (hard copies) to its editable form (soft copies) and they can be used in the future. This is more useful in Government offices where a number of documents have to be handled. In such offices maintaining a soft copy of the documents is more organized than keeping a hardcopy, especially for old documents. Therefore using this system we can keep soft copies of all these documents and hence the problem of damaged old documents can be avoided. In this project, we are trying to identify and convert old scripts of Malayalam language to its present age new scripts using machine learning. It involves the following steps. First the system is trained with different styles of Malayalam characters. Then image of the Malayalam old scripts is given as input to the system. Then noise is removed from the scanned image. Then the image is converted to text which is the new Malayalam script.

REFERENCES

[1] M Abdul Rahiman, M S Rajasree, "An Efficient Character Recognition System for Handwritten Malayalam Characters Based on Intensity Variations", International Journal Of Computer Theory And Engineering, Vol.3, No.3, 369 - 374, June 2011

[2] Masha N, Rema M, Meenakshi R, Manoj Kumar G, "Recognition of Handwritten Malayalam Characters using Vertical & Horizontal Line Positional Analyzer Algorithm", 2011 3rd International Conference On Machine Learning And Computing (ICMLC 2011)

[3]Jomy John, Pramod K.V, Kannan Balakrishnan, Bidyut B Chaudhuri, Department of Computer Applications, Cochin University of Science and Technology,Kochi, A two stage approach for handwritten Malayalam character recognition, IEEE 2014, 14th International Conference on Frontiers in Handwriting Recognition

[4]Gayathri P, Sonal Ayyappan, Department of Computer Science and Engineering,SCMS School of Engineering and Technology, Ernakulam, Off-line Handwritten Character Recognition using Hidden Markov Model, IEEE 2014, International Conference on Advances in Computing, Communications and Informatics (lCACCl)

[5]Steffy Maria Joseph, Dr Abdul Hameed ,Signal Processing and Communication Lab, Govt. Engineering College, Kozhikode, Online Handwritten Malayalam Character Recognition using LIBSVM in Matlab, IEEE 2014