Kannada Character Recognition and Translation to English

[1]Chandraprabha K S, [2]Aishwarya K Kamati, [3]Bharati B Akkalakumpi, [4]Chaithra D L

[1]Assistant Professor, [2,3,4]VIII Semester Students
Department of CSE
Siddaganga Institute of Technology, Tumakuru, India

Abstract: **As we all know that Kannada is the regional language since it is spoken only by Karnataka people in that also very minimum number of people using Kannada as their mother tongue, Most importantly we all know that Karnataka is having a very good climatic conditions like it has so many rivers, oceans, temples, different styles of tasty and healthy food, culture and most importantly green trees hence these things grouped together makes Karnataka map so attractive in such a way that many of the foreigners would like to visit holly and tourist places in Karnataka to make them fresh, when they came from that faraway places they are also curious about to know about the culture, tradition and history of that particular places which they visit since in Karnataka its practice to put the boards about speciality of that particular place to give more knowledge about that place hence it can be easily read by the people those who are rom Karnataka only but it's a difficult task for a people those who don't know Kannada but they are interested to know what is written over the boards hence in order to help those kind of people to get more knowledge about Karnataka and its places which they visit we got motivated to do this particular project.**

I. INTRODUCTION

India consists of eighteen official languages in that kannada is one among them,it is spoken by around 48 million people all over the country and it is considered as the official language particularly for the south Inidan state of Karnataka. The kannada alphabets were discovered during the emperors ruling time that is between 5th and 7th centuries A.D.Moreover Kannada script is different from all other scripts and it's not like any north languages and so in order to overcome this problem we are developing our project to understand the content which is written in kannada on some boards when someone visits the tourists places. kannada scripts are more complicated than English and other languages, but it doesn't have any upper or lower case distinction.

Since due to the complicated compound words present in kannada some kannada unknown people and non-Karnataka people may find difficulty in reading the scripts written in kannada when they visit the tourists places In our project we are developing an android application this application will helps to understand the scripts written in the kannada by translating it to English so that everyone can understand the content and gain the knowledge and also we are doing the translation using machine learning algorithm and checking the accuracy among the two working systems**.**

II. LITERATURE SURVEY

[1] Implementing Kannada Optical Character Recognition on the Android Operating System for Kannada Sign Boards. It is an application can be used for general purpose. By using this application user can read and understand the Kannada text written on the various sign boards. They have used some of the algorithms like kohonen neural network and Hilditch algorithm. Kohonen neural network algorithm for character recognition and Hilditch algorithm for image processing to remove noise and other disturbances and also to reduce the font size. Image processing is done to improve the recognition rate of the system. In this Optical character recognition is done with the help of Kohonen neural network. They have used neural network for better accuracy and the processing speed.The approach the have followed is first they have optimized the captured input image to get the best results and that is compared with the database. Characters from the image are recognised and equivalent English meaning is displayed along with the voice. In this work finding equivalent English meaning is not done by comparing it in the backend. Instead of that each Kannada words meaning is found by connecting to the internet and using popular Kannada dictionary sites. when the image is not clear or when the application is not able to recognise the text then the error message is displayed. If the internet is slow then also error message is displayed.

[3] Kannada Characters and Numerical Recognition System using Zone-Wise Feature Extraction and Fused Classifier. In this work they have used a hybrid zone based feature extraction algorithm and fused classification for handwritten offline Kannada character recognition. For feature extraction they have considered combination of both distance metric and pixel density algorithms. And also for fusion of classifiers both K-nearest neighbour and linear discriminant analysis is used. They have achieved good recognition rate by using fused classifiers. They have also showed that good recognition rate can be achieved by combining two or more classifiers rather than individual classifier. Image processing is done using the methods like thinning, binarization and size normalization. In this approach feature extraction is done using zone wise and classification is done by combining both KNN AND LDA. In feature extraction they have proposed algorithms like finding image centroid and zone centroid by using corresponding distances and the density of the pixels in the given image. In this work they have compared the results of individual KNN and LDA results with the fused classifier. As per the result the recognition rate is more In fused classifiers rather than individual classifiers. In this work they have developed both Kannada character and numerical recognition system. They have just recognised the Kannada characters and numbers and not translated to any other languages like English. Database they have considered is Kannada 49 alphabets which includes 13 vowels and 34 consonants. Initially they have considered only distance metric to extract features but accuracy was less so they have decided to use both pixel density and distance metric.

[3] Basic Kannada Handwritten Character Recognition System using Shape Based and Transform Domain Features.In this work they have created their own database by collecting various images consisting of Kannada vowels and consonants because standard

Kannada is not available. They have proposed an algorithm for recognition of Kannada handwritten vowels and consonants by using normalised chain code and wavelet transform based features. They have given more importance to the feature extraction step because the classification and recognition is dependent on the features. The feature extraction methods used are Normalised chain code and wavelet decomposition. All pre-processed images are normalised to particular size for measuring the performance of an algorithm. They have used Support vector machine(SVM) classifier. And the performance of the algorithm is tested using two fold cross validation method. The main aim of this proposed system is to remove the confusions when similar shape characters are found. The novelty of this system is free from the thinning. The images are cropped manually and that are converted into grey scale images. The proposed algorithm gives best results to meet the objectives of the Kannada OCR system to compared with the other existing methods.

III. SYSTEM DESIGN

System Architecture
The below figure represents the overall architecture of the proposed android application. In our proposed application user need to capture the image which consists of Kannada text. Image need to be imported from the gallery. The selected image is given to the android application which recognizes the Kannada text from the given image and translates the Kannada text into the English using Google cloud and translated text is displayed on the screen.
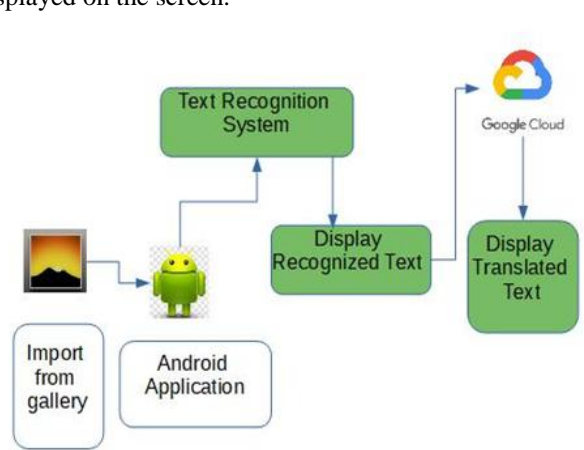

Figure 1.1 System Architecture

Activity Diagram
The below figure represents the activity diagram of the proposed application. I our proposed application user need to import the image consisting of kannada text. If image is not imported then the application stops, else image is the given to text recognition system. If internet is turned on the application translates the kannada text into english using Google API cloud and translated text is displayed else turn on the internet and continue the process.
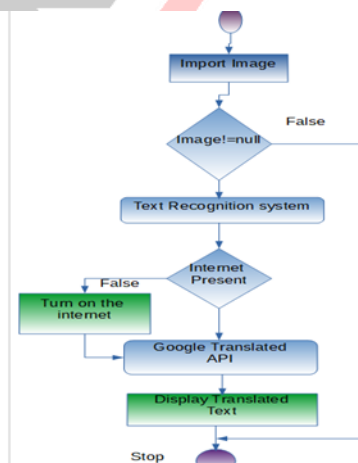

Figure 1.2 Activity Diagram

IV. TOOLS AND TECHNOLOGY

OCR
OCR represents Optical Character Recognition which is used to recognize the text contain in the image, for example scanned documents and photos. This technology is developed by Hewlett-Packard between 1985 and 1994.It is used to convert a kind of images containing written text to machine readable text data. These days solutions deliver near to perfect OCR accuracy. When a scanned paper document undergo OCR processing, the content of the document can be altered with word processors like Microsoft

Word or Google Docs. Earlier the only option is that digitize printed paper documents was to manually re-typing the text. It may leads with inaccuracy and typing errors. At the prior phase of presenting OCR, the technology has a few issues, for example, limitations in the quantity and complexity of the hardware and the algorithm. Be that as it may, OCR has been generally utilized in many areas including check processing, digital libraries, perceiving content from characteristic scenes, understanding handwritten office structures and so forth.

Tesseract Tess-Two Recognition Engine
Tesseract is an OCR engine which support for unicode and ability to recognize the many languages. It can be trained to recognize the other languages. It is an open source of OCR library. it can be integrated with Android application was developed by Hewlett Packard Labs and was then released as free software under the Apache license 2.0 in 2005.
We made sense of it would be more convenient if it was more portable,subsequently we chose to build up an Android Application for Kannada Language. Android gadgets accompany a camera which we can use to capture the image directly.

Implementation details
The working details of the proposed application is as follows:
1.User need to capture image which consists of Kannada text from the android phone's camera.
2.Import the image from the gallery and the same is given to text recognition system.
3.Tesseract OCR is used for Kannada text extraction
And the extracted text is translated to the English using Google Translate API
4.The translated English text is displayed on the screen along with the voice

V. CONCLUSION
The Kannada Character Recognition and Translation to English Project is designed to make it easier for people who is not known about Kannada when they visit tourist places it becomes difficult for them to understand when they visit holly places when something is written on boards so in order to overcome that we have developed this project.
Since India has many holly places in order to get some relaxation and peace many people plan there holidays to visit some holy places and also there is a myth that when we visit holly places and takes bath in sacred rivers then until that if we have committed any mistake then that sins will get rid off from us, by adding this this scenario and many reasons were there to visit holly places. So when we are planned to visit holly places we may find many boards from source to destination at that time while looking for our way we may find it difficult when we see Kannada words so at that time in order to get what is written on boards and helps to the people to find their way we have developed this application which takes input from the user as image, once after taking image as input internally extraction will happen there after it is stored in file which is abstracted from the common users so by clicking on update button they will get the content which is written on boards and get conveyed in a understandable manner.

REFERENCES
[1] Manjunath A E ,Sharath B," Implementing Kannada Optical Character Recognition on the Android Operating System for Kannada Sign Boards".
[2] Kavya.T. N, Pratibha.V, Priyadarshini. B. A, Vijaya Bharathi. M, Vijayalakshmi.G.V," Kannada Characters and Numerical Recognition System using Zone-Wise Feature Extraction and Fused Classifier".
[3] Shashikala Parameshwarappa, B.V.Dhandra," Basic Kannada Handwritten Character Recognition System using Shape Based and Transform Domain Features".
[4] Yalniz, I.Z.; Manmatha. R,"A Fast Alignment Scheme for Automatic OCR Evaluation of Books", International Conference Document Analysis and Recognition (ICDAR), 2011.