

International Conference Computational Intelligence & Data Science

February 4, 2021

East West College of Engineering,
Yelahanka New Town, Bengaluru



In collaboration with Universal Innovators

Principal
A.J. Institute of Engineering & Technology
Mangaluru - 575 096

International Conference Computational Intelligence & Data Science

February 4, 2021 | East West College of Engineering, Yelahanka New Town, Bengaluru

BACKGROUND

International Conference on Computational Intelligence and Data Science (ICCIDS-2021) is organized with the objective of bringing together scientists, professors, research scholars, students and industrial experts in the field of Data Science and its various Applications to a common forum. Overall, the conference will provide the researchers and attendees with prospects for national and international collaboration and networking. ICCIDS-2021 will be held at East West College of Engineering, Yelahanka New Town, Bangalore. All the accepted papers (after double blinded peer review) are published in Elsevier SSRN.

OBJECTIVE OF THE CONFERENCE

The primary objective of the conference is to provide opportunity for academicians, industry experts, practitioners, professionals, researchers and policy makers from different fields to engage in discussion based on issues related to dynamic and challenging economic environment. It will also provide a platform to get acquainted with latest developments and trends in the economy and business environment coupled with their implications for the organizations. The forum will facilitate interaction among members inside and outside their own respective disciplines to enjoy the fellowship of other professionals and scholars in the field.

FOCUS AREAS

- To facilitate discussion on Computational Intelligence & Data Science that will expedite nation's growth and stability.
- To bridge the gap between academic wisdom and practical knowledge.
- To develop a series of focused research topics that will benefit the academic world.

Director General Chair:



Dr. Santhosh Kumar G
B.E, M.Tech, Ph.D
Principal / Director

Chief Patron:



Dr. Purushotham B
M.Sc., M.Phil., Ph.D., PGDBA, PGDMB, B.Ed.
DIRECTOR - Admissions
EWGI

Patrons:



Smt. Rashmi Ravikiran
Chairman
East West Group of Institutions



Mr. Tejas Kiran C R
Secretary
East West Group of Institutions

CONFERENCE CHAIR:

Dr. Aravind K U
Professor & Head of Department – M.E
Member, RESEARCH AND INNOVATIVE CELL (RIC)
East West College of Engineering Yelahanka New Town, Bangalore

Convener:

Dr.Piyush Kumar Pareek,
Professor & Head of Department of Computer Science and Engineering
East West College of Engineering, Bengaluru

ADVISORY COMMITTEE

- Dr Ashish Khanna, Computer Science and Engineering Dept., Maharaja Agrasen Institute of Technology, GGSIPU, New Delhi
- Dr Deepak Gupta, Computer Science and Engineering Dept., Maharaja Agrasen Institute of Technology, GGSIPU, New Delhi
- Dr.E.Madhusudhana Reddy, Professor, Department of Computer Science and Engineering , Guru Nanak Institutions Technical Campus(AUTONOMOUS), Telangana, India.
- Dr Raghavi K Bhujang, Head of Computer Applications and Information Science, ISBR Business School, BLORE
- Dr. Manoj H M, CSE, Affiliation: BMS Institute of Technology and Management, Bangalore
- Dr Dinesha H A, Prof & Head, CSE, SGBIT, VTU, BELAGAVI
- Dr.C.P.S.Prakash, Department:Mechanical engineering,Affiliation: Principal, Dayananda Sagar College of Engineering
- Dr Rajnish Ratna Gedu college of Business Studies Royal University of Bhutan
- Dr. P Priyanga, Associate professor , Computer science and engineering , Global Academy of Technology, Bangalore
- Dr S.Venkatesan , Prof &Dean , School of Computer,Information &Mathematical Sciences , BSA Crescent institute of science and technology , Chennai
- Dr Latha p h, Hod dept of ISE,Vtu, Sait
- Dr R V Praveena Gowda, PhD , Principal,G Madegowda Institute of Technology, Karnataka
- Dr.Sunanda Dixit , Computer Science and Engineering Department,BMS Institute of Technology and Management, Bangalore
- Dr. Sonal Amit Jain , Computer Science Engineering, JK Lakshmipt University, Jaipur
- Dr.Sangeetha V , Information Science and Engineering , SVIT , Blore
- Prof. Chandrasekar Shastry,Dean PG Studies, Faculty of Engineering & Technology,Jain (Deemed to be) University,Bangalore, Karnataka.
- Prof.Santosh Kumar J , Associate Professor,Department of CSE,K S School of Engineering and Management , Holiday Village Road, Vajarahalli Village, Mallasandra, off, Kanakapura Rd, Bengaluru
- Dr. Raghavendra S , Associate Professor,Department of Computer Science and Engineering, School of Engineering and Technology, CHRIST DEEMED TO BE UNIVERSITY. BENGALURU

TECHNICAL PROGRAMMING COMMITTEE

- Prof. Nishanth.P,H.O.D,Department of Aeronautical Engineering,East West College of Engineering, Bangalore.
- Prof. Umapathi M A,H.O.D .Civil Engineering Department,East West College of Engineering,Bangalore.
- Prof. Rashmi H N,H.O.D,Department of Electronics and Communications Engineering,East West College of Engineering,Bangalore.
- Prof. Venkatesh G,H.O.D,Department of Basic Science and Humanities,East West College of Engineering,Bangalore.
- Dr.Geetha.N,Training and Placement Officer,East West College of Engineering,Bangalore.
- Dr. Babu N,Department of Mathematics,East West College of Engineering,Bangalore.
- Dr. Gufran Ahmad Ansari, Prof. & Head, Computer Applications, B.S. Abdur Rahman Crescent Institute of Science and Technology, Vandalur, Chennai
- Dr.Narasimha Murthy M S , CSE , EWIT,Blore

Principal

A.J. Institute of Engineering & Technology
Mangaluru - 575 006

CALL FOR PAPER AND SUGGGESTED THEMES

Original papers on the following indicative topics, but not limited to, are welcome to be submitted for the conference:

Artificial Intelligence and Machine Learning.
Data Mining and Text mining.
Big Data Analytics and Metrics.
Cloud Computing.
Mobile Computing.
Natural Language Processing and Machine Translation.
Parallel and Distributed Algorithms.
Pattern Recognition and Analysis.
Modeling Systems and Software Engineering
Ubiquitous and High-Performance Computing.
Information, network and applications security.
Access control, web filters and firewalls.
Intrusion detection and incident response.
Vulnerability assessment and mitigating attack.
Social engineering and phishing attacks
Information security management and governance.
Software security and secure coding.
Cognitive Radio and Cognitive Networks.
Communication Architecture.
Control Systems and Applications.
Real-Time Networking.

Satellite and Space Communications.
Smart Grid Systems and Intelligent Control.
Smart Communication System Models.
Industrial Networks and Automation.
Cognitive Approach for Robotics.
Mobile Robots and Intelligent Autonomous Systems.
Human-Machine Interfaces.
Image and Video processing.
Object recognition.
Speech processing.
Medical imaging.
Renewable Energy and Power System
Power Electronics
Electromagnetic and Remote Sensing
Integrated Systems, Circuits and VLSI Design
Nano Technology & Photonics
MEMS, NEMS
Systems, Controls
Robotics
HVAC and HVDC Systems

A C C E P T A N C E A N D P U B L I C A T I O N

After a double-blind peer review, qualifying Regular Papers may be accepted as either Full Papers or Short Papers.

- All accepted and presented papers of the conference will be included in ELSEVIER-SSRN digital library (ISSN 1556-5068).
- The papers must be part of the worldwide scholarly discourse in the field covered by the library. The reviews will be done to make sure papers are relevant for the chosen classifications to ensure subscribers receive relevant content.
- NO extra fee is charged from authors for inclusion in the digital/eLibrary. Most of the articles in the digital/eLibrary will have the full text of the article available online in open access so authors can view/download their papers at NO extra cost.
- NO subscription is required on the part of any organization, school, or individual author as a prerequisite to submission or download.
- Authors will grant a non-exclusive, revocable license that allows providing services to users.
- The proceedings page would be permanently available, not need to pay an additional fee to keep the papers available on the digital library.

I N D E X I N G

Post-conference, SSRN proceedings will be made available to the following indexing services for possible inclusion:

- ISI Conference Proceedings Citation Index - ISI Web of Science
- Google Scholar
- Scopus
- DBLP

Depending on the focus of the particular indexing services, they may decide to include or not. If included one can expect it in 10-12 months. DBLP and Google Scholar are fast.

*Some high-quality papers will be considered for Scopus Indexed Springer Proceedings (The selection of paper in Springer/SSRN will be made while sending the acceptance notification).

<https://online-conf.com/iccids-home.html>

Principal
A.J. Institute of Engineering & Technology
Mangaluru - 575 098

**3rd INTERNATIONAL CONFERENCE ON COMPUTATIONAL
INTELLIGENCE AND DATA SCIENCE
February 4, 2021|EAST WEST COLLEGE OF ENGINEERING**



ELSEVIER

SL.NO	TITLE	PAGE NO
1	ACCURATE FRAUD DETECTION IN CREDIT CARD TRANSACTIONS USING HYBRID HEURISTIC AND META-HEURISTIC ALGORITHMS	1
2	OPTIMAL WEB SERVICE COMPOSITION USING HYBRID OPTIMIZATION ALGORITHM IN CLOUD ENVIRONMENT	2
3	DEEP REINFORCEMENT LEARNING BASED INTELLIGENT TRAFFIC CONTROL SYSTEM	3
4	FAULT DETECTION OF BEARING USING XGBOOST ALGORITHM AND DATA VISUALIZATION USING T-DISTRIBUTED STOCHASTIC NEIGHBOR EMBEDDING (T-SNE) METHOD	4
5	DEVELOPMENT OF MULTISOURCE CLOUD LOGS FORENSIC WITH ENCASE TOOL FOR PERFORMING LOCK TECHNIQUE USING DEEP ANALYSIS AND REPORTING	5
6	ASSESSING BLOCKCHAIN IN SUPPLY CHAIN MANAGEMENT	6
7	PATIENT DATA DE-IDENTIFICATION IN CLINICAL RECORDS USING DEEP LEARNING	7
8	SECURED STORAGE PATH USING CRYPTOGRAPHIC VIRTUAL MAPPING FOR CLOUD APPLICATIONS	8
9	SENSITIVITY CONTEXT AWARENESS BASED PRIVACY PRESERVING RECOMMENDER SYSTEM	9
10	IMPROVEMENT OF POWER QUALITY ISSUES USING UNIFIED POWER QUALITY CONDITIONER IN DISTRIBUTION SYSTEM	10
11	ALTMETRICS ANALYSIS IN SOCIAL MEDIA USING BIGDATA	11
12	DATA ANALYTICS FOR BUSINESS USING TABLEAU	12
13	CUSTOMER CHURN PREDICTION IN TELECOMMUNICATION INDUSTRY THROUGH MACHINE LEARNING BASED FINE-TUNED XGBOOST ALGORITHM	13
14	PERFORMANCE EVALUATION OF BACK PROPAGATION ALGORITHM BY CHANGE IN LEARNING RATE AND MOMENTUM VALUE – A MACHINE LEARNING APPROACH.	14

15	SMART DRIP IRRIGATION SYSTEM USING IOT	15
16	VIEW INVARIANT GAIT AUTHENTICATION USING TRANSFER LEARNING	16
17	RENEWABLE ENERGY HARVESTING INVERTER USING SUPER-CAPACITOR	17
18	DETECTING DUPLICATE QUESTIONS IN COMMUNITY BASED WEBSITES USING MACHINE LEARNING	18
19	DETECTION AND LOCALIZATION OF MASK OCCLUDED FACES BY TRANSFER LEARNING USING FASTER RCNN	19
20	ROLE OF ITOPS IN DEVOPS	20
21	BLOCKCHAIN-BASED PREFERENTIAL E-VOTING SYSTEM DAPP USING SMART CONTRACT	21
22	AN APPROACH TO ANALYSIS ON COVID-19 DATA THROUGH WEB SCRAPER AND VOICE ASSISTANT – A SURVEY	22
23	WEB ACCESSIBILITY TEST APPROACH ON DEVELOPED WEBSITES	23
24	AN AMELIORATED METHODOLOGY TO DETERMINE RELATIONS AMONG OBJECTS	24
25	TEXT IDENTIFICATION SYSTEM FOR TRANSLATION OF ENGLISH LANGUAGE	25
26	DATABASE CREATION FOR NORMAL AND SUSPICIOUS BEHAVIOR IDENTIFICATION IN ATM VIDEO SURVEILLANCE	26
27	AUTHENTICATION OF A USER USING A COMBINATION OF HAND GESTURE AND ONLINE SIGNATURE	27


Principal

A.J. Institute of Engineering & Technology
Mangaluru - 575 006

3rd INTERNATIONAL CONFERENCE ON COMPUTATIONAL INTELLIGENCE AND
DATA SCIENCE

February 4, 2021|EAST WEST COLLEGE OF ENGINEERING



ELSEVIER

**TEXT IDENTIFICATION SYSTEM FOR TRANSLATION OF ENGLISH
LANGUAGE**

Mrs.Sushma¹,Mrs.Snitha Shetty²,Mrs.Aparna Shetty³,Mrs. Saranya Babu⁴ Mrs. Sharon D'souza⁴

Department of ISE, 2 Department of CSE, A J Institute of Engineering and Technology,

Mangaluru (KA)

sushma@ajiet.edu.in, snitha@ajiet.edu.in, aparna@ajiet.edu.in, saranyababu@ajiet.edu.in

sharon@ajiet.edu.in, putturpradeep@ajiet.edu.in

ABSTRACT :

India is a multilinguistic country. People of different states speak different languages but all Indians are not polyglots. English is called as universal language and Kannada is one of Dravidian languages that is used in India. The majority of people in Indian, especially who live in villages, cannot read, write and understand English language. Translation of any language that is done by human is time consuming. Therefore, there is a need for implementation of Machine translation systems that will translate English language to Kannada language. This paper describes translation system that enhances the knowledge of society without language barrier by translating text in English language to Kannada language by giving the meaning of the word. We can click the picture of any English text and give Image as input to translation system. The system translates English text to Kannada text by extracting characters from an image by using tesseract software. Tesseract is optical engine that scan image containing text and extract characters from image and convert them into editable form. Extracted text which is in English language might have non dictionary words that may or may not be translated properly to Kannada language by existing system. It also identifies such non dictionary words compares with hybrid database and replaces these words with exact Kannada word.

Text Identification System for Translation of English Language

Mrs. Sushma^{a1*}, Mrs. Snitha Shetty^{b2}, Mrs. Aparna Shetty^{b3}, Mrs. Saranya Babu^{b4} and Mrs. Sharon D'souza^{b5}

^a Department of Information Science and Engineering, AJ Institute of Engineering and Technology, Mangaluru, 575006, India

^b Department of Computer Science and Engineering, AJ Institute of Engineering and Technology, Mangaluru, 575006, India

¹ Sushma@ajiet.edu.in ² snitha@ajiet.edu.in ³ aparna@ajiet.edu.in ⁴ saranyababu@ajiet.edu.in ⁵ sharon@ajiet.edu.in

Abstract: India is a multilinguistic country. People of different states speak different languages but all Indians are not polyglots. English is called as universal language and Kannada is one of Dravidian languages that is used in India. The majority of people in India, especially who live in villages, cannot read, write and understand English language. Translation of any language that is done by human is time consuming. Therefore, there is a need for implementation of Machine translation systems that will translate English language to Kannada language. This paper describes translation system that enhances the knowledge of society without language barrier by translating text in English language to Kannada language by giving the meaning of the word. We can click the picture of any English text and give Image as input to translation system. The system translates English text to Kannada text by extracting characters from an image by using tesseract software. Tesseract is optical engine that scan image containing text and extract characters from image and convert them into editable form. Extracted text which is in English language might have non dictionary words that may or may not be translated properly to Kannada language by existing system. It also identifies such non dictionary words compares with hybrid database and replaces these words with exact Kannada word.

1. Introduction

Translation of one language to another language is very important in every day of a life to share information and communicate. Translator is used in businesses to meet market demands and legal requirements and it can also be used gain competitive advantage. Translators are used by travel industries to provide translation facilities to holiday makers. Due to large number of language, Machine translation has high significance in India. Kannada is a Dravidian language which is used by Forty three million people as their mother tongue. Availability of machine translators that translates English to Kannada language pair is low. The motivation behind English to Kannada machine translation system is to afford tranquil learning skill of innovative language. Many document and government records, historical records are written in English that is not understood by many people who live in village. Especially rural villager in Karnataka who know only their native language that is Kannada. Hence, translation system that translates English to Kannada language is needed to translate given text and to provide exact translation for non-dictionary words. For text extraction, image containing text is given as input. An image consists of smallest element called pixel. Each pixel in an image corresponds to one value. Pixel value at any point corresponds to intensity of light photons that is striking at that particular point. At particular location, value is stored by each pixel that is proportional to the intensity of light at that location. There are 5 main formats to store images that are explained below.

1.1 Tagged Image File Forma

The file type ending with extension .tif. File that has very large size is created by TIFF image. TIFF image contains a lot of detailed image data because it is uncompressed. It is flexible in terms of content and image. Photo software and page layout software uses this type of file because this file contains lot of data related to image. It is mainly used for printing a photograph. This files cannot be used in web because it

will take more time to load. TIFF cannot be used for web image because web browser cannot display TIFFs.

1.2 Joint Photographic Experts Group

The file type ending with extension .jpg. It is also called as JPG. It is an image that is compressed to store a large amount of information in file that is small size. Most of the image that is captured by digital camera is stored in JPEG format because more no of photos can be taken on single camera card than with other formats. Through compression jpeg misses specifics about the image to acquire a trifling file. Accordingly it is acknowledged as “loss” compression. JPEG files create a very small file which can be easily loaded to web page so these files can be used for web photographs. JPEG files look “bitmappy” because of compression so these files are not suitable for line drawings. JPEG can be used for web projects, documents of Microsoft office or in a project where high resolution printing is required. TIFF has disadvantages compared to JPEG that is memory card can store only few images and it takes longer time to transfer image to card. JPEG cannot be used for line art and on image which has sharp edges with uniform color, JPEG gives poor result.

1.3 Graphic Interchange Format

Graphic Interchange Format is the file type ending with .gif. In this format image is compressed but different from JPG, compression is lossless that is details of the image is not lost during compression. File in GIF are not as small as JPEG. GIFs is suitable for web but it is not suitable for printing and photography because of limited color range. It can be used for animations. GIF can be used if image has less number of color less than 256 and for the image which has areas of uniform color. File size of GIFF is drastically reduced due to limited number of colors.

1.4 Portable Network Graphics

Portable Network Graphics is file type ending with .png. Which was fashioned to replace gif with exposed setup since manifest for Portable Network Graphics was retained by Solitary Corporation and people were not ready to pay a fees regarding license. It allows colours of full range and provide better compression. It can be used for web images but not for print images. Portable Network Graphics is not as suitable as jpg, because larger file is created. It is better for the image with line art and some text because image is less “bitmappy”.File which is smaller in size is produced by PNG and it will allow more colours. It supports partial transparency. Transparency is used for antialiasing of text and fades.

Principal
A.J. Institute of Engineering & Technology
1.5 RAW Image Files
Mangaluru - 575 006

Usually it contains data from camera. Since these files are not processed it is called as raw files. Size of these files are very large

because raw files contains huge amount of uncompressed data. These files cannot be printed or edited.

1.6 Properties of Text

Text in image shows variations because of the difference in these properties listed below

Size: Text in an image may have various size.

Alignment: Text in an image can be aligned in different direction and caption text is aligned horizontally which will lead to geometric distortions.

Color: It is difficult to extract characters, if characters have similar color as the background.

Edge: Scene and caption text can be read easily since there is robust boundaries at border line of manuscript and circumstantial.

Compression: Images transmitted, processed plus recorded in compacted setup. Text abstraction is faster if extracted without decompression.

1.7 Types of Text

Image contains various types of text that are listed below

Document Text: Document text are the text that comes from papers, book and journal. Text in this document may vary in style, size, color and alignment. Extracting text from these document are difficult due to color of the background and foreground text.

Caption Text: It is overlaid on image or video and describes the theme of an image. These text that is overlaid might be influential basis of linguistic. These text has the application in sports for designing of sports highlights and definition.

Scene Text: It is the text that exists in captured video or image and it contain linguistics information like advertisement that contains name of shops, institutions and sign of roads. This is very useful in identification of vehicle license plate.

1.8 Text Extraction Process

Text extraction process takes input as image containing text and output the text data. It is castoff in numerous applications such as automobile authorization plate appreciation, involuntary bank crisscross processing, signboard detection and translation, text translation system, Content based image indexing, page segmentation etc. Text extraction process consists of four phases they are,

Text Detection: The Existence of text in image is determined.

Text Localization: This phase deals with the determination of the rigorous position of character in image besides outline nearby the character is marked.

Text Tracking: This phase will lessen the dispensation phase of video.

Text Extraction: Text component segmented from background. The Text Extraction system is shown in Figure 1.1

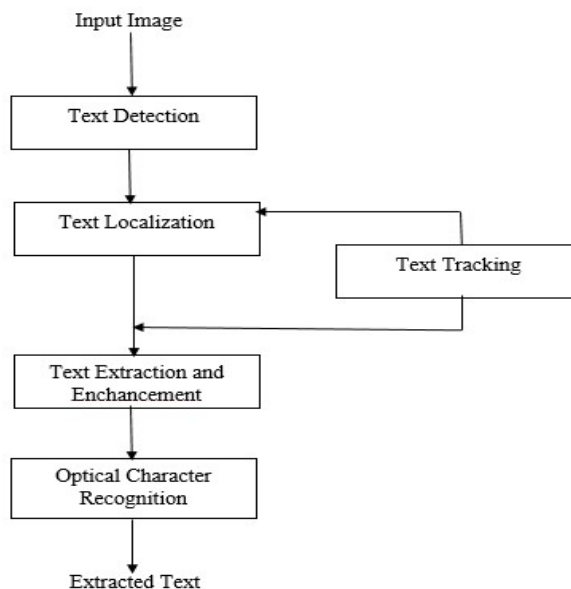


Figure 1.1: Architecture of Text Extraction

1.9 Machine Learning

Machine interpretation is procedure of transforming basis language to objective language with or without manual support. Basis language is the one that has to be translated. Target language is the one to which given source language has to be translated. Source language text exists is given as input to machine translator. Text which is translated to target language is generated by translator as output. Translation that is done manually is time consuming and it is expensive. Translation can be done quickly by using Machine translator and it will also reduce manual work. There are two types of machine translator bilingual and multilingual. Bilingual machine translator is used to translate from one source language to another target language whereas multilingual translators is used to translate set of languages. Several approaches used in translation are:

1.9.1 Rule Based Machine Translation

Based on grammatical, morphological, and lexical rule, Source text is parsed into an intermediate representation. These representation is converted into target language.

1.9.2 Direct Based Machine Translation

In this method, dictionary is used to translate individual word from one language to another. Word by word translation is another name for this method.

1.9.3 Corpus Based Machine Translation

CBMT based on the bilingual text corpora analysis. Two method comes under this method that is Example Based MT and Statistical MT .When translating English source text to Hindi target text Statistical MT is

used for catching exception. For translation, linguistic knowledge is not used for translation in SMT.

1.9.4 Knowledge Based Machine Translation

KBMT uses domain specific Knowledge Base. Based on semantic web and ontology the Knowledge Based is created.

1.9.5 Hybrid Based Machine Translation

Hybrid Based MT is the combination of any two or more machine translation techniques.

1.10 Optical Character Recognition

OCR is the software which takes documents such as PDF files, scanned document and images as the input and produces the editable text as output. The primary goal of OCR is digitizing scanned documents. Recognition of text is done automatically by OCR. Classical OCR systems can identify text with high accuracy and it can also maintain structural/layout information. This system require very high ratio of signal-to-noise and distortion-free orientation of text [1]. Few problems occurred in the development of OCR system that is visible difference between some digits and letters is very less for computer to understand. For example, it is difficult to differentiate between letter "o" and digit "0" for computer. Text which has very dark background and text that is printed on other words are difficult to extract. OCR application can be used for Authorization plate acknowledgement, extraction of character from ordinary section images, character extraction from perused documents. In this project ORC with tesseract is used for the extraction of text from an image. It takes input image and returns text file that contain text that is extracted from an input image in an editable form. Tesseract is open source and it can be compiled very easily on any platform.

2. Literature survey

Victor Fragoso, Jim Kleban, Steffen Gauglitz and Shane Zamora proposed method describes a multimodal translator which uses services of online translation and OCR and smartphone's camera and touchscreen is used to develop this translator to translate source language to target language [4]. Optical character recognition is the software that will scan images that contains texts and converts them into document that can be edited

C. Vasantha Lakshmi and C. Patvardhan proposed android mobile application that captures text from an image and translate it to telugu language and captured is saved in .jpg format. The text detection and extraction is done using tesseract OCR engine. For the purpose of learning, Translated text will be augmented on a screen of the mobile [3]. BehrangParhizkar, OneNdaba, KeamogetsweOteng, Zahra MohanaGebriel and ArashHabibiLashkari proposed method that is used build visual translator on mobile that translates Bahasa to english language for education purpose [2]. The user will have a compatible mobile phone application. Translation of any word can be generated by pointing a camera at any word. The word gets translated to another language. At same time, meaning of the word can be obtained.

Shibily Joseph and Nithya B proposed the English to Malayalam translator that uses hybrid approach. Hybrid approach is used here is the combination of SMT and TM [8]. Corpus is used as the primary requirement. This include both English Malayalam corpus and Malayalam corpus which is monolingual. Parallel corpus availability for Malayalam language is quite low. Hence proposed method uses training data set that is small corpus that has of 563 sentences. 150 sentences are used for tuning as tuning data set. Transliteration tools is used to accomplish translation.

Jayashree Nair, Deetha R and Amrutha Krishnan K proposed an English to Hindi translator. Declension rule is used to design this system. HMT architecture is proposed by this method that is used to split input sentence that is in English language to form set of words. These set of words are passed to parser. The Parser parses these set of words [7]. Analysing the semantics and syntax structure of sentences that are in English language is called as parsing. The word in English language is translated to Hindi word from lexical dictionary. The Proposed system can be improved to deal with complex sentences and it can also be enhanced to include a multiple languages.

3. Existing System

In this section, we describe previously existing system that can only translate few language to English language by extracting the text from the given image. Existing system may or may not translate the non-dictionary word correctly, if non dictionary word is given as input.

Limitations

User have to type the text that has to be translated to get the appropriate meaning of the text. It will not translates the English language to Kannada language by extracting the text from an image. Less accuracy is obtained due to poor lighting condition, blurriness and orientation problem. Input which is given to system might have non dictionary words that may or may not be translated properly to Kannada language.

4. Problem Statement

Since Kannada is the language of India that is used by many people. There are many people who can only read and write in Kannada language. There is no application to translate English language to Kannada language by extracting text from an image and there is no such application that provide exact translation for non-dictionary words. This application is used to translate the English language to Kannada language. If people of Remote villager who only know Kannada language and want to visit other tourist places, without translator it is not possible to communicate with the people of other Language and it is not possible for student to find the meaning of English word in Kannada language.

5. Proposed System

Proposed system will translates text that is in English language to Kannada language by extracting the text from an image which is given as input by user and it will also provide the meaning of given English text in Kannada language. Extracted text which is in English language might have non dictionary words that may or may not be translated properly to Kannada language.

6. Proposed System Architecture

The proposed English to Kannada translation system translates text in an English language to Kannada language by extracting text from an image and it also gives the meaning of the word that is translated. Architecture diagram of English to Kannada translation system is shown in Figure 6.1. Image is given as input to tesseract tool and this tool will extract text from an image and store extracted text in notepad file and these text is translated from English language to Kannada language. Extracted text which is in English language might have non dictionary words that may or may not be translated properly to Kannada language. Text in notepad file is splitted into number of words and check whether these words are available in English dictionary. If it is not available, then check if this non dictionary word exists in hybrid database that contains the key-value pair for representing English and Kannada word respectively. If word is found, then system will convert this character of word into lower case character and replace particular non dictionary Kannada words in translated Kannada text with exact Kannada words from the hybrid dictionary if the word is inappropriate else it will display as it is.

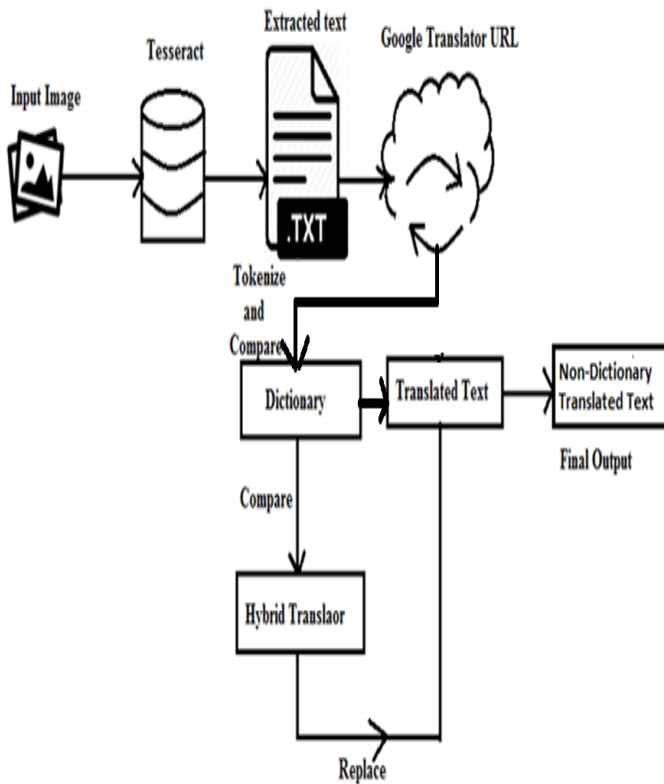


Figure 6.1: Architecture Diagram of Proposed System

6.1 Tesseract

Tesseract engine is the command based tool that is used to extract text in from image. Tesseract tool uses command that will take two arguments that are name of the file that contains text and output text file in which text that is extracted from image is stored. File extension .txt has to be specified for the output file name. Working of the tesseract OCR is shown below in Figure 6.2.

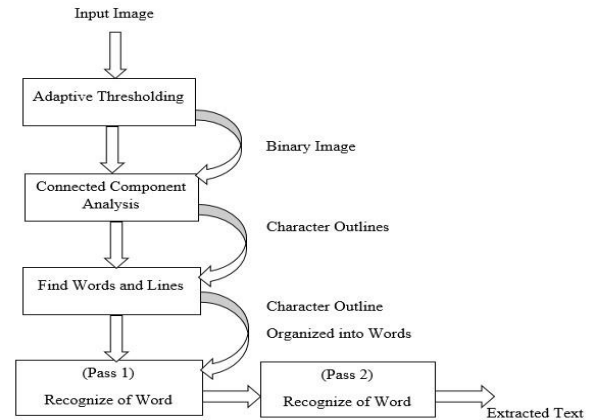


Figure 6.2: Tesseract Architecture

Adaptive Thresholding

It is the first step, in which the given image is converted into binary images. This converted binary image is given as input for connected component.

Connected Component Analysis

Binary image that is obtained from Adaptive Thresholding is analyzed to extract outline of the characters. This method is useful because it will do OCR of image with black background and white text. Then, these obtained outline of the character is converted into blobs. Blobs contains text lines and plain regions and blobs are analyzed for some equivalent text size.

Find Words and Lines

Blobs that is obtained from previous step it contains lines of text and plain region. Using fuzzy spaces and definite spaces, text is divided into number of words.

Recognition Word

Recognition of text is done in two-pass process. In first pass, each word is recognized. These words that are recognized is given to adaptive classifier which contains training facts. It will try to recognize word more accurately. Adaptive classifier has received training data that is something new. In final phase various issues are resolved and text is extracted from image.

6.2 Google Translator

The system uses Google Translator URL for translating the text from English language to Kannada Language. Block diagram is shown in Figure 6.3. In Google cloud platform, first step is to sign in with your Google mail account. Select project page and create new project by entering the project name. When project name is entered automatically

project id will be assigned to the project. Billing for the project is enabled by providing the details of card. Next step is to enable API for the project. Google API is obtained by enabling the credentials for the project. This particular API key is copied to and passed to URL.

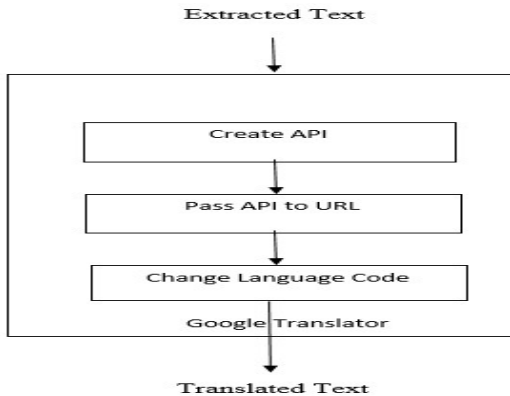


Figure 6.3: Block Diagram of Google Translator API

The source language and target languages are English and Kannada respectively. Code for Kannada language “kn” and English language is “en” is passed to URL for translation purpose.

6.3 Non-Dictionary Translation

Text extracted from the google translator may or may not contain non dictionary word. Filtering of these words are done by Splitting the extracted text from tesseract into number of words and then these English word is compared with the dictionary to obtain non dictionary word in extracted text and then it is compared with hybrid database. A hybrid data base that has key value pair for representing English and Kannada word. English word is considered as key word for which there exists a Kannada words. English word that exists in hybrid database is taken as a key and two Kannada words is taken as values. Next step is to check if there is any non-dictionary word that matches the word in hybrid dictionary. If word matches with hybrid dictionary word and has improper meaning in Kannada. Then, second Kannada word that has improper will be replace by first word which gives the proper meaning. Once the word is replaced the output is stored in notepad file.

7. Result and Analysis

The below section shows the screenshots of various inputs and outputs.

7.1 Notepad Image

The Figure 7.1 shows the input image created in Note pad.

Mangalore Express bus

Figure 7.1: Input Image

The output translation system is stored in text file for the given image shown in Figure 7.2.

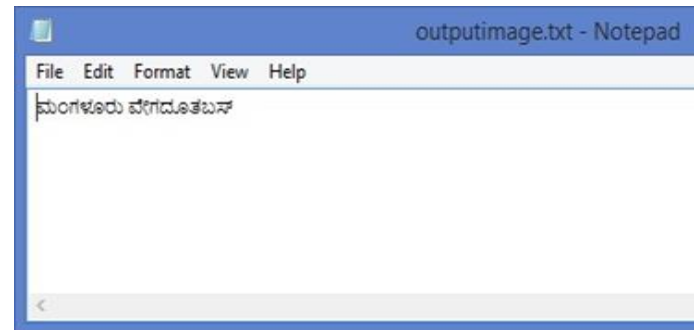
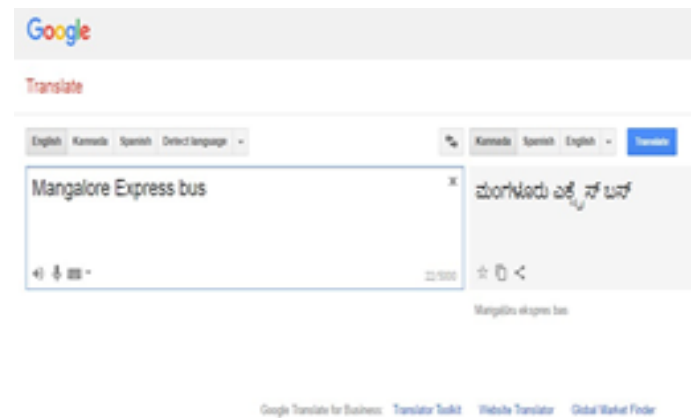


Figure 7.2: Output of Proposed System for Given Input Image

When this input image is passed to English to Kannada translation system it will translate the non-dictionary word “Express” correctly. Figure 7.3 shows the google translator output for sentence “Mangalore Express bus” in which the non-dictionary word “Express” is not



translated correctly.

Figure 7.3: Google Translator Output For Given Input Image

The Figure 7.4 shows the image containing the text “Ugadi celebration”.

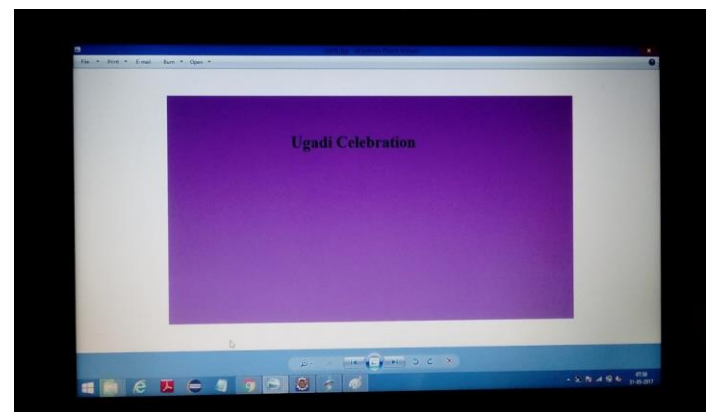


Figure 7.4: Camera Image

The output translation system is stored in text file for the image shown in Figure 7.5 translated the non-dictionary word “Ugadi celebration” correctly.

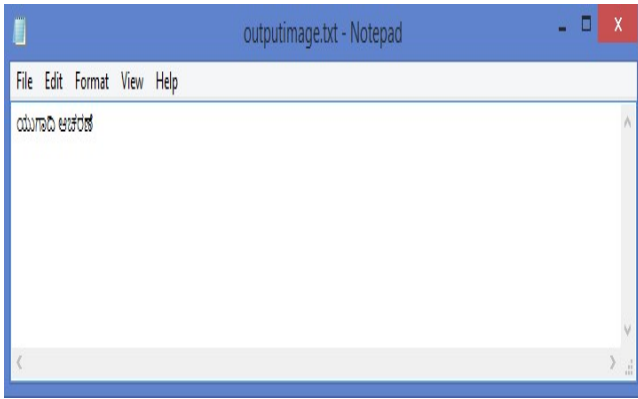


Figure 7.5: Output of Proposed System for Given Input Image

The Figure 7.6 shows the Google translator output for sentence “Ugadhi Celebration” in which the word Ugadhi is not translated correctly.

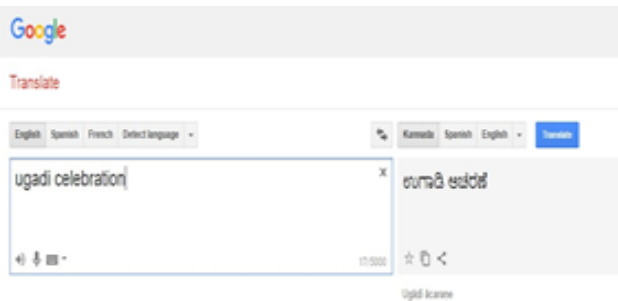


Figure 7.6: Google Translator Output For Given Input Image

7.2 Accuracy Analysis

The Figure 7.7 shows the text that is extracted from input image and the google translator output and output of the proposed English to Kannada machine translation system.

Text in Input Image	Google Translator Output	Proposed System Output
Ugadi Celebration	ಉಗಾದಿ ನೆಲೆವೇಡನ್	ಯುಗಾದಿ ಆಚರಣೆ
Explore beautiful mangaladevi temple location	ಸುಂದರ ಮಂಗಳಾದೇವಿ ದೇವಾಲಯ ಸ್ಥಳವನ್ನು ಅನ್ವೇಷಿಸಿ	ಸುಂದರ ಮಂಗಳಾದೇವಿ ದೇವಸ್ಥಾನ ಸ್ಥಳ ಅನ್ವೇಷಿಸಿ
Ramaa wore traditional saree	ರಾಮ ಸಾಂಪ್ರದಾಯಿಕ ಸೀರೆ ಧರಿಸಿದ್ದರು	ರಮಾ ಸಾಂಪ್ರದಾಯಿಕ ಸೀರೆ ಧರಿಸಿದ್ದರು
Soap will cause harm to eye	ಸೋಪ್ ರಚ್ಚಿಗೆ ಹಾನಿ ಕಾರಣವಾಗುತ್ತದೆ	ಸಾಬೂನುರಚ್ಚಿಗೆ ಹಾನಿ ಕಾರಣವಾಗುತ್ತದೆ
Mangalore Express bus	ಮಂಗಳೂರು ಎಕ್ಸ್‌ಪ್ರೆಸ್ ಬಸ್	ಮಂಗಳೂರು ವೇಗದೂತಬಸ್

Figure 7.7: Google Translator Output For Given Input Image

8. Conclusion

This paper explains the English to Kannada translation system that will extract the text which is English language from a given image and translate it to Kannada language by giving the meaning of the word. Extracted text which is in English language might have non dictionary words that may or may not be translated properly to Kannada language. The proposed system identifies such non dictionary words, compares with hybrid database and replaces these words with exact Kannada words if it is inappropriate. Application is useful for the people of remote village who can only read and write Kannada language. Application can be used by the individuals who want to learn Kannada language.

8. Future Work

This system can be enhanced by adding more words to the hybrid database and In future we can also add a feature that will translate Kannada text to Kannada speech and it can also be enhanced as mobile application.

REFERENCES

Faustina Jeya Rose.R and Bhuvanewar. *Word Recognition Incorporating Augmented Reality for Linguistic E-Conversion*. International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), 2016.

Behrang Parhizkar, Keamogetswe Oteng, One Ndaba, Arash Habibi Lashkari and Zahra Mohana Gebril. *Ubiquitous Mobile Real Time Visual Translator Using Augmented Reality for Bahasa Language*. International Journal of Information and Education Technology, Vol.3, No. 2, April 2013.

C.Vasanth Lakshmi, C. Patvardhan. *An optical character recognition system for printed Telugu text*. Pattern Analysis & Applications table of contents archive Volume 7 Issue 2, July 2004.

Victor Frago and Steffen Gauglitz. *TranslatAR: A Mobile Augmented Reality Translator on the Nokia N900*. Published in Applications of Computer Vision(WACV), 2011 IEEE Workshop on 5-7 Jan. 2011, pages 497 – 502.

Q. Yuan, C. L. Tan ,”Text Extraction from. *Gray Scale Document Images Using Edge Information*. Dept. of Computer Science, School Of Computing National University of Singapore 3 Science Drive 2, Singapore 117543.

Jayashree Nair, Amrutha Krishnan K and DeethaR. *An Efficient English to Hindi Machine Translation System Using Hybrid Mechanism*. 2016 Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI), Sept. 21-24, 2016, Jaipur, India.

Jayashree Nair, Amrutha Krishnan K and DeethaR. *An Efficient English to Hindi Machine Translation System Using Hybrid Mechanism*. 2016 Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI), Sept. 21-24, 2016, Jaipur, India. **Mangaluru - 575 006**

Nithya B,Thrissur. *A Hybrid Approach to English to Malayalam Machine Translation*. International Journal of Computer Applications (0975 – 8887) Volume 81 – No.8, November 2013

