

Literature Survey on Word Identification System for Kannada Language

Sushma¹ and Madesha M²

⁴th Sem M. Tech, Dept of CS&E, Sahyadri College of Engineering and Management, Adyar, Mangalore, India¹

Assistant Professor, Dept of CS&E, Sahyadri College of Engineering and Management, Adyar, Mangalore, India²

Abstract: Translating text from one language to another by extracting text from an image will enhance the knowledge of society without any language barrier. Text extraction from image involves detecting the text from given image, finding the presence of text location, extraction, enhancement and recognition of text from the given image. Machine translation technique is used to translate source language to target language. Machine translation plays an important role in benefiting linguists, sociologists, computer scientists. The demand of translation has become more in recent years due to increase in the exchange of information between various regions using different regional languages. A large number of techniques have been proposed to address this problem. The purpose of this paper is to review and classify various text extraction algorithms and translation techniques.

Keywords: Machine Translation, Text Location, Text Enhancement, Linguists, Sociologists.

I. INTRODUCTION

Text Extraction from image is concerned with extracting the text from the images. It has many useful applications such as Automatic bank check processing, vehicle license plate recognition, document analysis and page segmentation, signboard detection and translation, content based image indexing, assistance to visually impaired persons, text translation system for foreigners etc.

There are 5 main formats in which images is stored:

TIFF: TIFF stands for Tagged Image File Format. TIFF images create very large file sizes. TIFF Images are uncompressed and thus contain a lot of detailed image data. TIFFs are also extremely flexible in terms of colour and content.

JPEG: JPEG stands for Joint Photographic Experts Group. JPEG files are images that have been compressed to store a lot of information in a small-size file. Most digital cameras store photos in JPEG format, because then you can take more photos on one camera card than with other formats.

GIF: GIF stands for Graphic Interchange Format. This format compresses images but, as different from JPEG, the compression is lossless. GIFs also have an extremely limited colour range suitable for the web but not for printing. This format is never used for photography, because of the limited number of colours. GIFs can also be used for animations.

PNG: PNG stands for Portable Network Graphics. It also allows for a full range of colour and better compression. It's used almost exclusively for web images, never for print images.

Raw image files: Raw image files contain data from a digital camera. The files are called raw because they haven't been processed and therefore can't be edited or printed. There are a lot of different raw formats. Each camera company often has its own proprietary format.

The image content is classified into two categories: perceptual content and semantic content. Perceptual contents include colours, shapes, textures, intensities, and their temporal changes. While semantic contents include objects, events, and their relations. Text content contains high level of semantic information as compared to visual information. Text in images can exhibit many properties [11]:

Size: The text can be of variable size.

Alignment: The caption texts appear in clusters and usually lie horizontally. This does not apply to scene text, which has various perspective distortions. Scene text can be aligned in any direction.

Edge: An edge in the images is the most reliable feature of text regardless of intensity, orientations, etc.

Colour: In a simple image the characters in a text usually have the same colours. This property makes it possible to use a connected component-based approach for text detection.

Text extraction process mainly consists of four important phases that is [10]:

Text Detection: Presence of text in each frame or image is determined.

Text Localization: The exact location of the text is decided and boxed around them.



Text Tracking: The time of processing for localization of video is reduced and the integrity of position of texts across the adjacent frames of video is maintained.

Text Extraction: The segmentation of text component is performed from background.

A Text Information Extraction system receives an input as image and output the relevant text data. Images can be in grey scale or colour. The Text Information Extraction as shown in Figure1.

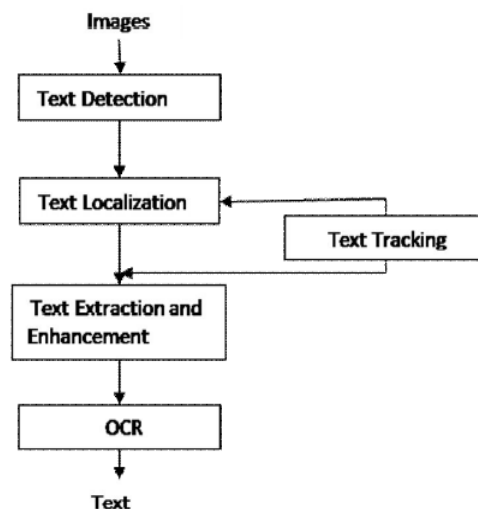


Figure1: Architecture of Text Extraction System

Machine Translation is the use of computers to automate some or all of the process of translating from one language to another. A machine translator usually deals with two languages namely a source language and a target language. Source language refers to the language that is to be translated and target language refers to the language to which the source language is translated. Thus the input to a translator is some source language text and the translator generates the equivalent target language text as the output. Machine translators can be bilingual or multilingual. A bilingual machine translator converts from one language to another whereas multilingual translators can translate languages from a specified set of languages to another set of languages. The machine translation systems have several approaches for translation[14]:

Rule Based MT: RBMT system parses the source text into an intermediate representation based on certain grammatical, lexical and morphological rules. The intermediate representation is then converted to the target languages.

Direct Based MT: Direct translation, also called as word by word translation, translates individual word in a sentence from one language to another using dictionary.

Corpus Based MT: CBMT depends on the analysis of bilingual text corpora. Statistical MT and Example Based MT comes under the Corpus Based MT. SMT is good for catching exceptions to rules when translating the English source sentence to Hindi target sentence. The benefit of SMT is linguistic knowledge is not required for translation.

Knowledge Based MT: A domain specific Knowledge Base is used in KBMT for translation. The KB has to be created based on ontology and semantic web.

Hybrid Based MT: HBMT is the combination of two or more MT techniques.

II. LITERATURE SURVEY

Victor Fragoso, Steffen Gauglitz, Jim Kleban and Shane Zamora proposed method that describes a multimodal, augmented reality translator developed using a smartphone's camera and touchscreen combined with OCR and online translation services. This paper presents a system implemented on a Nokia N900 phone which allows the user to simply hold up the phone, and, with a single click, have text translated into his/her language of choice and see the translation through the phone's display. System using standard of-the-shelf libraries which support multiple languages. The optical character recognition comes from an open source project called Tesseract. Finally, the rectified image data would be saved as a TIFF image and Tesseract would be used to retrieve its containing text. The translation is provided by Google's Translate web service. Figure 2 shows Augmented Reality Translator that detects and translates text area. Overlay the obtained translation onto the live video.

Major problems are especially non-uniform background and unusual fonts, text proportions, spacing, or sizes. While accuracy improves if the user manually annotates the region where the text lies in the scene. Result of application is



inaccurate for large fonts. It depends on real world imaging problems such as illumination variance, glare, and dirt are found to cause problems [4].

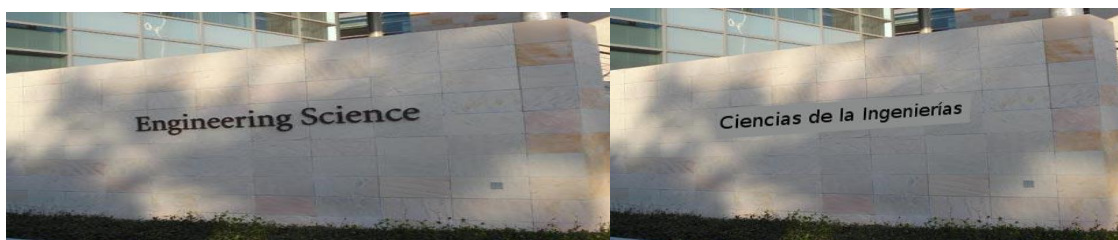


Figure 2: Augmented Reality Translator

C. Vasantha Lakshmi C. Patvardhan proposed android mobile application that translates English text to Telugu text. The process begins capturing text in an image and it is saved in .jpg format. The text detection and extraction is done using tesseract OCR engine [3]. Tesseract includes font properties files in assets folder of the project to identify the English text. TTF is the text file that contains font properties. Text which is extracted from OCR engine is given to Google translation engine as input for translation. Translation API checks database for available patterns, statistical machine translation is done on language, which is sent back to mobile screen. OCR produces accurate result if there is clear contrast between text and its background. Even the best available OCR engine fails to detect & extract text from the image that is captured from a portable mobile phone camera which may lead to low quality images. If lighting condition is poor or uneven the performance decreases. Accuracy of this method is highly influenced by orientation of camera. 100% accuracy in Text recognition is obtained when the lighting conditions are even. For future enhancement the application may be improved to include many such Indian languages.

Behrang Parhizkar, Keamogetswe Oteng, One Ndaba, Arash Habibi Lashkari, and Zahra Mohana Gebriel proposed method to combine real time visual with augmented reality technology on mobiles for education purpose [2]. The research is based on how the implementation of real time visual translator on augmented reality mobile applications can help people in understanding and give meanings to some writings displayed around them ubiquitously. The idea is to develop a translator that will translate Bahasa language to English language and give meanings of the words translated. The proposed application captures the words in Bahasa language and translates them to English language. On the other hand a directory/dictionary like Wikipedia, Oxford dictionary and so on, will be embedded on to the application so that after the application has finished translating the directory/dictionary will generate meaning of that word. Figure 3 shows a mobile phone displaying the word "sekolah" in Bahasa language and, on the other side it shows the mobile phone pointing at the converted word "school".

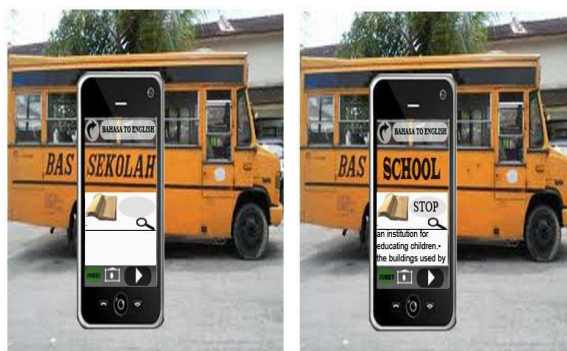


Figure 3: Translator Displaying on Mobile.

Q. Yuan and C. L. Tan presented a well-designed method that makes use of edge information to extract textual blocks from grey scale document images. It aims at detecting textual regions on heavy noise infected newspaper images and separate them from graphical regions. One most important step of the algorithm is to find approximate locations of text lines in a grey-scale image. If we compute the spatial variance along each horizontal line over the whole image, the regions with high variance correspond to text lines, and regions with low variance correspond to background. Moreover, through run-length smearing operation edges from textual regions may elongate to form the approximate encircling rectangles. Text lines can then be found by extracting the rows between two parallel edges of the spatial variance. This heuristic can be applied to a more complex image, assuming that the spatial variance in the background is lower than in the text. Text lines can then be found. One of the major weaknesses of our algorithm reside on its assumption that all text is



oriented in the same direction, which is by default horizontal. This makes the algorithm not suitable to deal with documents with multiple layout styles. From experiments we found the ratio of successful detection dropped under cases when textual paragraphs. Other than that, there also exist demands to increase the accuracy ratio on handling textual block extraction with complex background [5].

Xiaoqing Liu and JagathSamarabandu proposed a multistate edge-based text extraction algorithm, which can automatically detect and extract text in complex images[7]. The proposed method consists of three stages that is candidate text region detection, text region localization and character extraction. First stage aims to build a feature map by using three important properties of edges: edge strength, density and variance of orientations. The feature map is a grey-scale image with the same size of the input image, where the pixel intensity represents the possibility of text and global threshold can be employed to highlight those with high text possibility regions resulting in a binary image. Morphological dilation operator with a 7×7 square structuring element is applied to the previous obtained binary image to get text blobs. Four pairs of coordinates of the boundary boxes are determined. The purpose of final stage is to extract accurate binary characters from the localized text regions so that we can use the existing OCR directly for recognition to extract text from image. The proposed method can handle both printed document and scene text images and it is not sensitive to image colour, robust with respect to font, sizes, orientations, alignment, uneven illumination, and perspective and reflection effects. The proposed method only analyses text blocks. Therefore, it is computationally efficient, which is essential for real-time applications

Mohammad Khodadadi and AlirezaBehrad proposed algorithm for text detection, text extraction and in painting text in colour images. The aim of the text extraction algorithm is to extract text area from the candidate blocks. The text in each block has the same colour. The aim of the proposed algorithm is to inpaint the text area. For this purpose, we estimate histogram of colour channels for the text and background areas in order to extract the text characters in the candidate blocks[6].The proposed text extraction algorithm starts by the estimation of the histograms of colour channels for text and background areas. To estimate histogram for background, two areas with the height of two pixels above and below the candidate block is considered. Then the histogram of image pixels in these areas is calculated. The histogram curve is smoothed to remove noisy peaks. To estimate histogram of colour channels for text areas, we first extract pixels in the boundary of text characters. For this purpose, we employ threshold stroke image and determine pixels in the candidate block with high intensity change. Then these pixels are utilized to estimate the histogram for colour channels. Therefore, we compare the obtained histogram with the histogram of background to obtain text. This algorithm works well in text areas with medium or high contrast. This algorithm may fail when the text and background have very similar colours.

Oleg Okun, Yu Yan and MattiPietik"ainen proposed a method that is used to detect text in binary image using image filtering. Each pixel in the original image is replaced by rth pixel of its neighbours and new image is obtained [8]. The AND operation is performed on original image and obtained new image. Character shapes are preserved as much as possible by removing Black or white pixel that is isolated. By using this character shape text is detected and extracted from the image. This proposed method can be used to detect both inverted and normal text. It can be robust to different font styles and size. Chirag Patel, Dharmendra Patel and Atul Patel proposed method used to develop optical character recognition system called tesseract to extract text from the image. Tesseract engine is command based tool which takes image as input and extracts the text. Tesseract command takes two arguments: First argument is image file name that contains text and second argument is output text file in which, extracted text is stored. The output file extension is given as .txt by Tesseract[9]. Outlines of the text region are converted into Blobs. Blobs are organized into text lines, and the lines and regions. Recognition of text is then started as two-pass process. In the first pass, an attempt is made to recognize each word from the text. Each word passed satisfactory is passed to an adaptive classifier as training data. As adaptive classifier has received some training data it has learn something new so final phase is used to resolve various issues and to extract text from images. The results obtained by extracting vehicle number from vehicle number plate.

Mallamma V Reddy and Dr. M. Hanumanthappa proposed method for text pre-processing, translation of English to Kannada and then generating and implementing Phrase Structure Grammar for Kannada sentences. Part-Of-Speech tagger is the generator for assigning proper tags to each and every word in the training and test sentences. The proposed POS tagger used for parsing Kannada sentences and is implemented using supervised machine learning approach. Text pre-processing can be done by Lexical analysis that categorizes a text into digits, hyphens, case of letters and punctuation marks. The objective of filtering out words with very low discrimination values for retrieval purposes is possible by Elimination of stop words. Stemming of the remaining words with the objective of removing affixes and allowing the retrieval of documents containing syntactic variations of query terms. The suffix stripping and Suffix joining algorithm is implemented to make the stems as root words. Selection of index terms to determine which words/stems will be used as an indexing element. Usually, a particular word become an index term, the decision of it is depends on syntactic nature of the word. The frequency of noun words has more semantics than adjectives, adverbs, and verbs. The categorization structures construction defined as extraction of structure directly represented in the text, for allowing the extension of the original query with related terms[13]. Figure 4 shows the English sentence "I am fine" is converted to Kannada sentence "ನಾನುಚೆನ್ನಾಗಿಲ್ಲೆನೆ".

The Kannada Meaning for Sentence: iam fine

ನಾನು NN ಚೆನ್ನಾಗಿದ್ದೇನೆ VB

Back

Figure 4: Result of Translator

Jayashree Nair, Amrutha Krishnan K, Deetha proposed an HMT Architecture by the combination of declension Ruled Based MT and SMT systems. This system uses Splitter component to splits the inputted English sentences to a set of words. This set is passed on to the Stanford parser to analyse the syntax and semantics structure of the English sentence. The input sentence is passed through Stanford parser which checks the grammatical structure of each sentence. Parts Of Speech tagging for the parsed input source sentence can also be done using Stanford POS tagger. The POS tagger generates the information about the input sentences, words by suffixing alphabet based on the tag set. Tagging is done by loading a trained file that contains the information for the tagger to tag the input words. The tagged words are then passed on to the declension tagger that takes the tagged output from the POS tagger. The translator translates the English words to correct Hindi words from the lexical dictionary. Morphological analyser is used to analyse and transform each Hindi word based on its internal structure. This system design can be extend to deals with the complex compound sentences for translation. The system can also be used to develop a multi-language translation system for another Indian languages [14].

III. CONCLUSION

A large number of techniques have been proposed in the past but the detection of scene text and for translation of one language to another. These text extraction technique depends on additional complexities such as varying lighting, variable font sizes, style, colour, variance of orientation and complex background. The purpose of our paper is to classify and review various algorithms, discuss comparison and performance evaluation and to point out challenges for future research.

REFERENCES

- [1] FaustinaJeyaRose.R and Bhuvanewar," Word Recognition Incorporating Augmented Reality for Linguistic E-Conversion", International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), 2016.
- [2] BehrangParhizkar, KeamogetsweOteng, One Ndaba, ArashHabibiLashkari, and Zahra Mohana Gebriel," 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.
- [3] C.Vasantha 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.
- [4] Victor Fragoso and Steffen Gauglitz," TranslatAR: A Mobile Augmented RealityTranslator on the Nokia N900", Published in Applications of Computer Vision (WACV), 2011 IEEE Workshop on 5-7 Jan. 2011, pages 497 – 502.
- [5] 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.
- [6] Mohammad Khodadadi and AlirezaBehrad, "Text localization, extraction and inpainting in color images", Proceedings of 20th Conference on Electrical Engineering (ICEE), pp. 1035-1040, 2012.
- [7] Xiaoqing Liu and JagathSamarabandu, "Multiscale edge-based text extraction from complex images", 2006 IEEE International Conference, pp. 1721-1724, 2006.
- [8] Oleg Okun, Yu Yan and MattiPietikAinen "Robust Text Detection from BinarizedDocument Images",Machine Vision Group, Finland Received InRevised Form 12 October 2004, Accepted 14 January, 2005.
- [9] Chirag Patel, Atul Patel and DharmendraPatel,"Optical Character Recognition by Open Source OCR Tool Tesseract: A Case Study", International Journal of ComputerApplications Volume 55– No.10, October 2012.
- [10] Priyanka, DeelipWagh and D. R. Patil,"Text Detection, Extraction and Removal: A Survey", International Journal of Computer Applications National Conference on Emerging Trends in Computer Technology 19.
- [11]Aaditi Samadhiya, Pallavi Khatri, "General Review Of Various Text Extraction Techniques", India International Conference on Electrical, Electronics, Computer Science and Mathematics Physical Education &Management, (ICEECMPE), 8thNovember 2014, New Delhi, India, ISBN: 978-93-84209- 63-6.
- [12] VivekDhanapalSapate," A Survey: Text Extraction from Images and Video", International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 2, February 2016 Copyright to IJARCCCE DOI 10.17148/IJARCCCE.2016.5286 393.
- [13] Mallamma V Reddy and Dr. M. Hanumanthappa," POS Tagger for Kannada Sentence Translation",International Journal of Emerging Trends & Technology in Computer Science Volume 1, Issue 1, May- June 2012.
- [14] Jayashree Nair, Amrutha Krishnan K andDeethaR,"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.