Journal Homepage: - www.journalijar.com
INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)
Article DOI: 10.21474/IJAR01/3500
DOI URL: http://dx.doi.org/10.21474/IJAR01/3500
ISSN NO. 2320-5407

RESEARCH ARTICLE

SURVEY PAPER ON TEXT RETRIEVAL AND TRANSLATION FROM SIGN BOARD IMAGES.

Harshal Thosare, Mayank Pandey, Omkar Godse, Priyanka Shah and Prof. Mrs. Dipalee Choudhari.

Manuscript Info

Manuscript History

Received: 09 January 2017
Final Accepted: 04 February 2017
Published: March 2017

Key words:-
MSER, OCR, Text Detection, Text Retrieval, Text Translation

Abstract

The system assists people specially travelers to find their paths while travelling in English speaking countries without language barrier. Sign Board helps travelers to complete their journey with ease. But, understanding sign boards in English will be difficult for travelers who don't know English. To solve this, we are using text retrieval and translation from English to Marathi from sign boards. Text retrieval can be performed by using methods like Optical Character Recognition, Maximally Stable Extremal Region etc. Text translation can be performed by using various free API services available from organizations like Google, Microsoft, Yandex etc. For text retrieval, we found Optical Character Recognition as the best technique and for translation API services available from Google are the best.

Copy Right, IJAR, 2017,. All rights reserved.

Introduction:-

Sign boards provide route guidance using short and precise information. It conveys information that helps common people to take decisions while travelling. The language barrier among different English speaking countries is one of the major difficulties faced while travelling. The reason being tourists are unable to understand sign boards in the native language of the region. They may suffer consequent loss of important information which may lead to misguidance. The sign boards may also convey some warnings or hazards. The gaining momentum of internet and portable mobile phones makes it easy to drift from traditional dictionary translation to a better web app which is user friendly. The world is moving towards image records and retrieval of information from the same. In this sense, it is expected that a high percentage of the world population will own a smart phone with an embedded camera and a internet connection, which is our system's requirement.

Various techniques are used to retrieve text from an image. One of them is Optical Character Recognition (OCR)[1]. OCR is the process of converting a scanned document image into machine encoded format. The scanned document image may include photograph of a document, image of sign boards and billboard from a landscape photograph. OCR finds its applications in the fields of pattern recognition, artificial intelligence, handwriting recognition, etc. Although OCR works efficiently with digital images, it doesn't work well with natural scene images [2].

Another image processing technique is Maximum Stable Extremal Region (MSER). It is a technique used to detect blobs in images. Blobs are regions which differ in properties like brightness or color. MSER is useful while detecting text from a natural scene image. It also provides great results for images which are captured blur or of varying region density and size.

For taking pictures of the sign boards, we are using a mobile phone camera of a good quality so that we can get clear and stable pictures of the sign boards. Quality of a camera is going to affect the quality of the image of a sign board.

Corresponding Author:- Harshal Thosare.

So, we should take care of that while selecting the camera. Each of the methods of a text retrieval has some advantages and also some drawbacks. The next section i.e. Section 2 gives us the detailed idea about these methodologies used to retrieve the text and translate the text from sign board images. Section 3 gives us comparison between these methodologies. Section 4 gives us the conclusion of this survey.

Related Work:-
The system of text retrieval and translation from sign board images works in two phases i.e. text retrieval and text translation.

Text retrieval:-
It is the process of detecting & retrieving of the text from the images. There are two popular methods of text retrieval.

OCR- optical character recognition:-
The Optical Character Recognition(OCR) is considered as machine replication of human reading. Today's most common way of inputting the data into computer is by using the keyboard but, it is time consuming and it requires human efforts [3]. OCR provides a desirable way to extract the text from the images. So, that we can efficiently input the data to the computer. OCR is considered as mechanical or electronic conversion of scanned image where image can contain handwritten or typewritten text. OCR is the process of digitizing the printed text so that it can be used electronically in various machine processes. OCR is the process of converting a scanned document image into machine encoded format. The scanned document image may include photograph of a document, image of sign boards and billboard from a landscape photograph. OCR finds its applications in the fields of pattern recognition, artificial inteligence, handwriting recognition, etc.

Nowadays, many institutions and organizations are moving towards OCR to eliminate the human interference and increase the efficiency [5]. OCR is of two types and they are offline recognition and online recognition. In offline recognition either an image or scanned form of image is considered. In online recognition, successive points are chosen as function of time and order of strokes are also available. The OCR methodology has the following features:

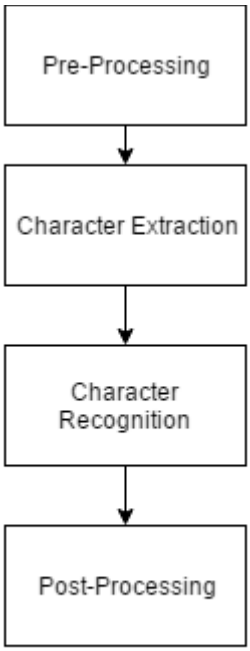
- No more retyping
- Edit text.
- Save space.
- Quick digital searches

Steps of a OCR are as follows:-
Pre-processing:-
OCR software often "pre-processes" images to improve the chances of successful recognition. Here input image is taken and it is converted to the gray scale image. The gray scale image is finally converted into binary image. This process is known as binarization of the image. Scanned image from the scanner may contain some noise due to the unnecessary details present in the image. This noise should be removed from the image for the proper application of OCR. Some algorithms are used to denoise the image and this denoised image should be referred for the further use in the system. Other operations that are performed in the pre-processing of the image are: Deskew, Despeckle, Fragmentation, Line removal, layout analysis.

Character Extraction:-
Preprocessed image is provided as input to this step to extract each and every character from text which is obtained from the given image.

Character Recognition:-
The image from the extraction stage is correlated with the templates which are preloaded into the system. Once the correlation is done, the template with maximum correlated value is declared as the character present in that image. There are two basic types of core OCR algorithm, which may produce a ranked list of candidate characters. Matrix matching involves comparing an image to a stored glyph on a pixel-by-pixel basis; it is also known as "pattern matching" or "pattern recognition". This relies on the input glyph being correctly isolated from the rest of the image, and on the stored glyph being in a similar font and at the same scale. This technique works best with typewritten text and does not work well when new fonts are encountered. This is the technique the early physical photocell-based OCR implemented, rather directly.

Post Processing:-
After the character recognition, if there are still unrecognized characters in the image then that characters are given their meaning in post processing step. Some extra templates can be added for providing wide range of compatibility checking in the system database. The Levenshtein Distance algorithm has also been used in OCR post-processing to further optimize results from an OCR API.



Block Diagram of OCR

We can face some challenges while implementing the traditional OCR on mobile devices. These challenges are-
1) Non-contact nature of digital cameras in mobiles.
2) Poor quality of images.
3) Processing speed and memory are not that sufficient to run the desktop OCR algorithms on mobile devices.
4) Mobile devices don't have Floating Point Unit(FPU) which is required for floating point arithmetic operations.
Hence, we are using computationally efficient and light weight OCR in our system which will run smoothly on the mobile devices.

MSER- Maximally Stable Extremal Regions:-
MSER denotes a set of distinguished regions, which are defined by extremal property of its intensity function in the region and on its outer boundary [10]. In computer science, Maximally Stable Extremal Regions(MSER) are used for blob detection in scanned images. This technique was suggested by Matas et al. to find correspondences between image elements from two images with different perspectives. This method of extracting a ample number of analogous image elements contributes to the wide-baseline matching, and it has led to better stereo matching and object recognition algorithms. Extremal regions in this context have two important properties, they are as follows:
1) Continuous transformation of image coordinates. This means it is affine invariant and it doesn't matter if the image is warped or skewed.
2) Monotonic transformation of image intensities. The approach is of course sensitive to natural lighting effects as change of day light or moving shadows.

Advantages of MSER are-
1) Invariance to affine transformation of image intensities.
2) Stability: only regions whose support is nearly the same over a range of thresholds is selected.
3) The set of all extremal regions can be computed in worst-case O(n), where n is the number of pixels in the image.

For region detection invariance transformations that should be considered are illumination changes, translation, rotation, scale and full affine transform (i.e. region should correspond to same preimage for different viewpoints. View point changes can be locally approximated by affine transform if assuming locally planar object and orthographic camera, that is perspective effects ignored). Region detection should be repeatable, stable and capable of discriminating between regions.

MSER is the method of blob detection in images. The MSER algorithm extracts from an image a number of co-variant regions called MSER. MSER is a stable connected component of gray level set of images. MSER is based on the idea of taking regions which stay nearly the same through a wide range of thresholds. All the pixels below a threshold are white pixels and all those above or equal are black pixels. If we are shown a sequence of threshold images $I_t$with frame t corresponding to threshold t, we would see first a *black* image, then *white* spots corresponding to local intensity will appear then grow larger. These *white* spots will eventually merge, until the whole image is *white*. The set of all related components in the order is the set of all extremal regions. Elliptical frames are attached to the MSERs by fitting ellipses to the regions. Those regions descriptors are saved as features. The word *extremal* refers to the property that all pixels inside the MSER have either higher(bright extremal region) or lower(dark extremal region) intensity than all the pixels on its outer boundary.

The operation can be accomplished by sorting all pixels by gray value and incrementally adding pixels to each connected component as the threshold is changed. The area is monitored. Regions such that their variation w.r.t the threshold is minimal are defined maximally stable.:

Let's make all the pixels below a threshold, white and all others, black. Considering a sequence of threshold images with increasing thresholds sweeping from black to white we pass from a black image to images where white blobs appear and grow larger by integration, up to the ultimate image. Over a enormous range of thresholds the native binarization is steady and shows some invariance to affine transformation of image intensities and scaling.

Text Translation:-
Our system is mainly divided into two parts, first one is text detection phase and second one is text translation phase. In text detection phase, we are using OCR and MSER together. In text translation phase, we are using text translation API. There are various free APIs available in market for the translation purpose like Microsoft Translator API, Google Translator API, Yandex Translator API, Frengley Translator API etc. Out of these APIs, some APIs are free and some APIs are paid. We have used the free API for the translation purpose in our System.

Microsoft Translator API:-
Microsoft Translator Text API [13], part of the Microsoft Cognitive Services API collection, is a cloud-based machine translation service supporting various multiple languages. This Translator can be used to build various applications, websites, tools, or any solution requiring multilanguage support.
It is Built for business. Microsoft Translator is a proven, customizable, and scalable technology for machine translation. Microsoft Translator technology powers translation features across Microsoft products, including Office, SharePoint, Yammer, Visual Studio, Bing, and Skype. By simply integrating translation into web, desktop, or mobile applications, the Translator API provides a rich functionality set for any developer. The features of Microsoft Translator API are- it is built for enterprise, customizable, effectively scalable.

Yandex Translator API:-
The Yandex Translator API [12] gives access to the Yandex online machine translation service. It provides support for more than 70 languages and can translate separate words or complete texts. The API makes it possible to use Yandex Translate in a mobile apps or web services for end users Or to translate large quantities of text, such as technical documentation.We can use any of the following interfaces for accessing the Yandex Translate API over HTTPS:

- XML interface (The response is an XML document).
- JSON interface (The response is JavaScript objects with the same names and semantics as the XML elements).
- JSONP interface (The response as JavaScript objects wrapped in a callback function with the specified name).

All the interfaces have the same functionality and use the same set of input parameters.
Yandex Translate is a statistical machine translation system. This system translates words, complete texts, and web pages into the desired language. It is available as a web service and mobile application, and is also used in other

products, such as translating web pages in Browser. Yandex Translate has an automated dictionary that sets it apart from the limited number of similar existing services. The technology, developed by a Yandex team of linguists and programmers, combines current statistical machine translation approaches with traditional linguistic tools.

Py-translate:-

Py-translate is a translation tool for humans. Its end goal is to translate text from one language to another. Translation is done in terminal using command line arguments. It is so good that it can translate a book in mere 5 seconds. It was made for python 3 but still it works on python2.It is fast & easy to install. It is easy to use & handle. It supports translation from any language. It is having highly composite interface, the power of Unix pipes & filters. It is having a simple API.

Comparison:-

Comparison of OCR and MSER

	OCR	MSER
Font Size	Fixed	Variable
Robustness	Less	More
Usability	Easy	Hard
Reliability	More	Less
Flexibility	Less	More
Time Complexity	Fast	Slow
Space Complexity	More	More

Comparison of Google API, Microsoft API and Yandex API

	Google Translate API	Microsoft Translator API	Yandex Translate API
Supported Request Format	Atom, GData	SOAP, XML, URI Query String/CRUD, JavaScript, HTTP POST, GET	XML, JSON, JSONP
Supported Response Format	GData, Atom, XML	XML, JSON, SOAP	XML, JSON
Is this Hypermedia API?	No	Yes	No
Languages Supported	100+	60	46
Functionality	Cloud Based API	Cloud Based API	Statistical Machine Translation

Conclusion:-

This paper focuses on the various methodologies that are used for text retrieval and translation from the images of the sign boards. By referring to the above sections, we can conclude that for the text retrieval, Optical Character Recognition(OCR) is better than any other methods for the digital images & Maximally Stable Extremal Regions(MSER) is best for natural scene images. As they produce the desired output with high accuracy. For the text translation, API services which are available from the Google are the best.

Future work aims at the development of a such system that performs text retrieval by using above methods and translation of that text from one language to another for the sign board images.

References:-

1. Shalin Chopra, Amit Ghadage, Onkar Padwal, Karan Punjabi, "Optical Character Recognition", IJARCCE, Vol.2, Issue 1, January 2014..
2. Pranob K Charles, V.Harish, M.Swathi, CH.Deepthi. "A Review on the various Techniques used for Optical Character Recognition", International Journal of Engineering Research and Applications(IJERA), Vol.2, Issue 1, Jan-Feb 2012.
3. Udo Miletzki, "Character Recognition in practice Today and Tomorrow", 1996,, Siemens Electrocom GmbH D-78767 Konstanz, Germany.
4. "Combination Of document Image Binarization Techniques", 2011 International Conference on Document Analysis and Recognition.
5. Ayatullah Faruk Mollah, Nabamita Majumder, Subhadip Basu, Mita Nasipuri, "Design of an Optical Character Recognition System for Camera based handheld Devices", IJCSI International Journal of Computer Science Issues, Vol.8, Issue 4, No 1, July 2011.
6. M.D. Dunlop, S.A. Brewster, "The Challenge of Mobile Devices for Human Computer Interaction", Personal and Ubiquitous Computing, 6(4), pp.235-236, 2002.
7. Mikael Laine, Ollo S. Nevalainen, "A standalone OCR system for mobile camera-phones", Personal, Indoor and Mobile Radio Communications, 2006 IEEE 17th International Symposium, pp.1-5, Sept.2006
8. Xi-Ping Luo, Jun Li, Li-Xin Zhen, "Design and implementation of a card reader based on build-in camera", International Conference on pattern Recognition, 2004, pp.417-420.
9. Xi-Ping Luo, Li-Xin Zhen, Gang Peng, Jun Li, Bai-Hua Xiao, "Camera based mixed-lingual card reader for mobile device",, International Confernce on Document Analysis and Recognition, 2005, pp. 665-669.
10. Ron Kimmel, Cuiping Zhang, Alexander M. Bronstein, Michael M. Bronstein, "Are MSER Features Really Interesting?", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 33, 2011.
11. J. Sivic and A. Zisserman, "Video Google: A Text Retrieval Approach to Object Matching in videos," IEEE Int'l Conf. Computer Vision, 2003
12. https://www.microsoft.com/en-us/translator/translatorapi.aspx
13. https://tech.yandex.com/translate/
14. https://pypi.python.org/pypi/translate