TEXT AND AUDIO TRANSLATION OF TEXT FROM SIGNBOARD IMAGES - REVIEW

Amit B. Kore[1] and D. R. Mehta[2]
[1]Student, [2]Associate Prof.
Department of Electrical Engineering, Mumbai University, Matunga, Mumbai

ABSTRACT
Text detection and recognition in images taken in uncontrollable environments is a challenging task. Extraction of information from a given image involves detection, localization, enhancement and recognition of the text. Variations of text due to differences in size, style, orientation, and alignment, as well as low image contrast and complex background make the problem of automatic text extraction extremely difficult and challenging job. A variety of techniques have been proposed to address this problem. Purpose of this paper is to review a whole architecture for translation of text present on scene text images like signboard images and to design a useful, simple, affordable robust system for text extraction which can be implemented easily using image processors and to identify promising directions for future research.

KEYWORDS: *Discrete Cosine Transform (DCT), Text extraction, Text Detection, Text Translation.*

I. INTRODUCTION

Google Goggles and World Lens are one of the notable text translation applications. However they still have certain limitation therefore it is important for image processors to be able to select a signboard and extract a character sequence from images. Signboard information plays an important role in our society. Their format is often concise and direct, and the information they give is usually very useful However, the foreign visitors may not understand the language that the signboard is written in, with the consequently loss of all that important information.

The application scenario considered in this paper is as follows. A user uses a camera to capture an image. By applying image pre-processing techniques the text area is to be extracted for recognition and finally it can be applied to applications to get the further text information. The next step is to translate the extracted text to Spanish language both in text as well as in audio format.

This method takes into account some basic problems during the text extraction like low resolution and low quality images, uneven lighting from shadowing, reflections.

This paper is organized as follows: Section II gives a brief description of background of text in images and information about different stages in general present in OCR System. Text area extraction method is proposed in section III. Section IV deals with Recognition and Translation of the extracted text to other language. Section V presents the step by step performance evaluation for various stages of OCR system. Section VI represents the conclusion of the study. Finally future work will be referred to in section VII.

II. TEXT IN IMAGES

Text in images is mainly classified into caption text which is artificially overlaid on the image and scene text which exists naturally in the image. Since scene text can have any orientation and it also

gets affected by different camera parameters such as illumination, focus, motion of camera. Hence it is difficult to detect scene text in the image.
A scene text occurs naturally as a part of the scene and contains important semantic information such as advertisements that include artistic fonts, names of streets, institutes, shops, road signs, traffic information and board signs (Fig.1).[2]

(a) (b)

Figure 1: (a) Scene Text Image, (b) Caption Text Image

2.1. Properties of Text in Images

Texts usually have different appearance changes like font, size, style, orientation, alignment, texture, color, and background. Text in images can exhibit variations with respect to following properties [2].

2.1.1. Size and alignment

Text size can very to great extent in the same image so it is necessary that the system should be able to handle fonts of different sizes. In the scene text characters can have various perspective directions.

2.1.2. Color

The characters on signboard images are supposed to be easily recognizable so they have strong contrast with background.

2.1.3. Edge of Image

Most of the caption and scene text are designed to be easily read, thereby resulting in strong edges at the boundaries of text and background.

2.2. Text Information Extraction (TIE)

The TIE problem can be divided into the following sub-problems: (i) Detection, (ii) Localization, (iii) Extraction and (iv) Recognition (OCR) as depicted in Fig. 2

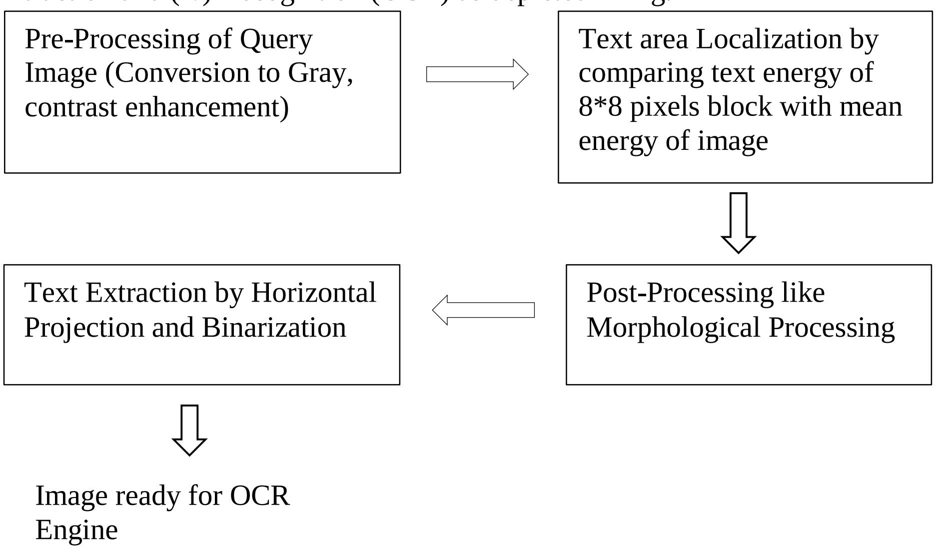

Figure 2: Architecture of TIE System.

2.2.1. Text Detection

It refers to the determination of the presence of text in a given query image.

Figure 3: Query Image to Tesseract OCR

The output of Tesseract OCR is
‘;;°ON THE smear
WHAT HAPPENS
** %ON THE STREET . A*
STAYS

When we input the image (Figure 3) without any pre-processing to OCR engine like Tesseract OCR[11] the results we obtained were not acceptable considering that the complexity of text extraction in those images was not very high. In order to improve the result obtained by Tesseract OCR we need to apply certain pre-processing steps to the query image which includes adaptive thresholding, morphological processing etc.

2.2.2. Text localization

It is the process of determining the location of text in the image and generating bounding boxes around the text. Various image properties such as energy; entropy etc. can be used to determine whether the pixel belongs to background part or text part. Image feature types can be divided into 2 groups: statistical features and structural features.

Statistical features are numerical measures of image like energy, entropy, chain codes, mean, variance, moments. Main advantage of statistical feature is they are easy to calculate and compare to check the similarity between two images or two regions of images.

Structural features mainly concerns with geometrical feature of image like end points, intersection points, perimeter of contour.

Since energy of text image differs largely from background image especially in signboard images where contrast between background and text part is higher. So we are using energy of block of 8*8 pixels in image to determine the text area.

2.2.3. Text tracking

Extracted text image after this process has to be converted to a binary image and enhanced before it is fed into an OCR engine (Tesseract v1.03).Tesseract OCR can only deal with image with simple, standard text. We often take photos by smartphone and there are number of environmental and camera parameters that affect the quality of the captured image. Text feature extraction is the stage where the text components are segmented from the background. Text feature is parameter by which the text object is uniquely represented from the image.

2.3. The Survey of Text Extraction Using Different Approaches

Table 1: Survey of Text Extraction using different approaches

Technique	Author	year	Title	Method
Region-Based	Leon [3]	2009	Caption text extraction for indexing purposes using a hierarchical region-based image model.	hierarchical region-based
	Debapratim[4]	2009	A Bottom-Up Approach of Line Segmentation from Handwritten Text	Bottom-Up Approach of Line Segmentation
	Chaddha [5]	1995	Text Segmentation in Mixed Mode Images	bottom-up region growing approach with Heuristic Filtering
	Karin[6]	2002	Identification of Text on Colored Book and Journal Covers	Clustering, Top-down successive splitting, bottom-up region growing
Morphological Based	Rama Mohan[7]	2010	Text Extraction From Hetrogenous Images Using Mathematical Morphology	Thresholding using morphological operators
	Jui-Chen Wu[8]	2008	Morphology-based text line extraction	A novel set of morphological operations & an x-projection technique
	Yuming [10]	2008	Text String Extraction from Scene Image Based on Edge Feature and Morphology	Mathematical Morphology and Edge border ratio

2.4. Scope and Discussion

Several techniques have been developed for extracting the text from an image. The proposed methods are based on morphological operators, wavelet transform, artificial neural network, edge detection algorithm, histogram projection technique etc. In this paper text extraction algorithm is proposed which can be applied to luminance images and which uses fast, simple and accurate arithmetic since the computation resources are limited.

III. TEXT CANDIDATE REGION EXTRACTION TECHNIQUE

3.1. Region -Based Technique

Region-based methods use the properties of the color or gray-scale in a text region or their differences with the corresponding properties of the background. This method uses a bottom-up approach by grouping small components into successively larger components until all regions are identified in the image.

Chaddha proposed an algorithm to differentiate text and non-text region in image [5]. In this method image is split in 8*8 pixels block and Discrete Cosine Transform is performed individually on each block. Energy of each block is calculated by empirically determining DCT coefficients to be 3, 4, 5, 11, 12, 13, 19, 20, 21, 43, 44, 45, 51, 52, 53, 59, 60, 61. Heuristic filtering is used to further increase the accuracy of text area detected. Similar methodology is proposed in this paper to detect and extract text regions from images. Only DCT coefficients selected to calculate the text energy are changed.

3.1.1. Text Candidate spotting

To remove and suppress the constant background Discrete Cosine Transform (DCT) based high pass filter is used. Text regions in the image are detected using the frequency information of the DCT (Discrete Cosine Transform) 8x8 blocks of the JPEG picture. DCT separate the low-frequency changes from the high-frequency changes. Since the eye can't detect high-frequency patterns well, these are quantized more heavily (i.e. with fewer bits) than the low frequency ones. The principle advantage of image DCT transformation is the removal of redundancy between neighbouring pixels. This leads to uncorrelated transform coefficients. The other advantage of DCT is it concentrates energy into lower order coefficients as compared to the DFT for image data. Hence DCT Transform on image is used [1].

3.1.2. Text characteristics verification

Kohei et.al [10] has introduced an approach to detect and extract text from images. He implemented edge based method and connected component analysis known as blob extraction method with appropriate thresholding. An easy but effective approach is presented in this paper for text candidate region determination using Discrete Cosine Transform of image. First element of DCT matrix represents a little information about the image because it is DC coefficient of image. To reduce the computation we will calculate sum of absolute values of the coefficients a_{12}, a_{13}, a_{14}, a_{15} of each block as other values are approximately zero. Now calculate mean text energy of the image. The block contains text if its Text Energy > 1.45*(Mean text Energy). [1] Next step is to create binary matrix such that Value 1 indicates block contain text and Value 0 indicates block contains background part.

3.1.3. Consistency analysis for output

Binary matrix is modified to obtain a more useful character representation as input for an OCR. The final step of Consistency analysis for output is performed by a Binarization algorithm that robustly estimates the thresholds on the caption text area [3].

3.2. Morphological Based Technique

Morphological analysis is a powerful tool for extracting geometrical structures and representing shapes present in an image. Main advantage of morphological analysis is the feature is invariant against various geometrical image changes like translation, rotation, and scaling. [7] Even after the lighting condition or text color is changed, the feature still can be maintained.

Morphological based methodology is proposed in this paper to reduce the effects caused by environmental and camera parameters like out of focus, lighting distortion and foreground-background color variations. Morphological closing followed by an opening operation is applied to this binary matrix with the structuring element [1 1 1 1], so that the text block candidates are merged together[1]. To further increase the efficiency we can apply edge detection using canny edge detector.

IV. TEXT BINARIZATION AND RECOGNITION

4.1. Text Binarization

Due to different variations in the image binary text matrix may indicates some part of the background as a text region. To avoid this we have to apply certain methods such as Thresholding, Binarization on the extracted text matrix. Binarization is the process on which the text area is divided in two different groups: foreground and background. Instead of applying Binarization algorithm directly to whole image it is proposed that we first divide the image in several parts and apply the binarization algorithm independently to each part, where the background and foreground colors are not supposed to change that much. By filtering the image through canny edge detector and then analysing the horizontal histogram of the edge image the text image is split into several parts. After this apply binarization algorithm independently. We can also apply color clustering algorithm and differentiate foreground from background by calculating Manhattan distance between two groups. The two groups (x, y) that maximize the separation between themselves in terms of the Manhattan distance (Equation 1) are probably the foreground and the background. [1]

$$d_m\,(x,y) = \sum_{i=1}^{n} |x_i - y_i| \qquad (1)$$

$$f(x, y) = d_m(x, y) \qquad (2)$$

4.2. Text Recognition

Now we have binary matrix in which text region is represented by 1 and background region is represented by 0. This binary image is applied as an input to an open source OCR (Optical Character Recognition) algorithm with well-known reliability: Hacking Tesseract v1.03. We can also use VietOCR .NET a GUI front end for Tesseract OCR engine for recognition of text from binary image. In VietOCR .NET [11] we give binary image in tiff format as a query image (figure 3) and then we do OCR processing on that image which gives us the text present in an image (figure 4) in the form of ASCII text. We can save the recognized text in separate file in txt format.

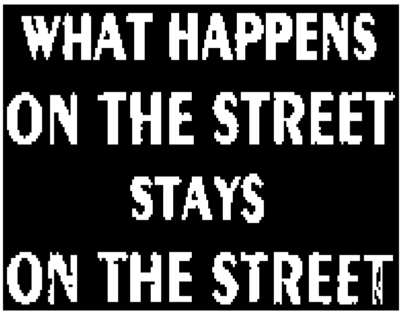

Figure 4: Pre-processed Binary Query Image

Figure 5: Output of Tesseract OCR Engine

From Fig. 5 we can conclude that after applying pre-processing technique to the query image (Fig. 3) and converting it to binary image (Fig. 4) we get good results.

4.3. Text Translation

One of the widely used methods for offline translation of text is phonetic transcription method [1]. It is based on a maximum likelihood criterion: Each letter is changed into the Spanish letter or letters that better approach its English sound. The results may not be accurate in all the possible cases. Advantage of phonetic transcription is the user will know how to pronounce the text extracted from image but in order to get good results system should be programmed with a big set of phonetic transcription rules. The other disadvantage of this system is if we want to change the English text to other regional languages like Japanese, Chinese, German, French, Hindi, Bengali etc. then considerable changes need to be done. Nowadays internet connectivity is easily available in the mobile and since our aim is to develop a simple and fast text recognition and translation algorithm which can be implemented in mobile phones we can use the Google Translate API for text translation as well as text to speech (tts) translation. With Google Translate API we can translate English text to any other language as per our requirement in text as well as audio format.

V. PERFORMANCE EVALUATION

Table 2: Processing Time comparison for text extraction system.

	Matlab 7.6.0 (R2008a) in a Pentium 4 PC (CPU 3.8 GHz, 2.99 GB RAM)	Matlab 8.1.0.604 (R2013a) in a Core-i3 PC (CPU 2.2 GHz, 2 GB RAM)
Reading of Query Image	0.27 seconds	0.09 seconds
Text area detection steps	0.29 seconds	0.32 seconds
Text information Extraction	1.43 seconds	1.34 seconds
Recognition of text	0.35 seconds	0.30 seconds
Translation of recognized Text	0.18 seconds	0.23 seconds

Recognized Text to Speech Conversion	NA	0.96 seconds
Total Processing Time	2.52 seconds (No voice Conversion)	3.24 seconds

VI. CONCLUSION AND DISCUSSION

The results of the experiment showed that the total processing time for text extraction system is acceptable but we should try to minimize it further. Processing time is not always constant and it largely depends on the size of query image.

This algorithm also doesn't work satisfactorily for texture background images.

Every approach has its own benefits and restrictions. Even though there are many numbers of algorithms, there is no single unified approach that fits for all the applications. The detection of small font text in presence of large font text also affects the text extraction efficiency of OCR system. It is difficult to extract small font information in such cases. Horizontal projection approach often gives inefficient results for small fonts. Heuristic Filtering approach along with connected component analysis is another approach to separate the text individually [5].

Conversion of translated text to voice will help user in knowing how to pronounce the text also. This is easy to understand as compared to reading the translated text.

We can implement this OCR architecture in a vehicle information system, for safety improvement and road convenience can be supplemented with a translation system for foreigners who do not understand the language written on signboard images

VII. FUTURE WORK

The future studies mainly concentrates on developing an algorithm for extraction of text from images which can work satisfactorily for several kinds of images like caption text, scene text, document text images and translation should be done as per the required language and to design the text area extraction system by SVM and HMM[12], and then design recognizer system.

REFERENCES

[1]. Canedo-Rodriguez, A. Kim, J.H., Soohyung Kim and Blanco-Fernandez, Y. (2009) "*English to Spanish translation of signboard images from mobile phone camera*," IEEE Southeastcon , pp. 356 – 361 DOI 10.1109/SECON.2009.5174105.

[2]. C.P. Sumathi, T. Santhanam, and G.Gayathri Devi, (2012)"*TECHNIQUES AND CHALLENGES OF AUTOMATIC TEXT EXTRACTION IN COMPLEX IMAGES: A SURVEY*". Journal of Theoretical and Applied Information Technology, Vol. 35, No.2.

[3]. Leon, M., Vilaplana, V., Gasull, A. and Marques, F., (2009) "*Caption text extraction for indexing purposes using a hierarchical region-based image model*," IEEE ICIP 2009, El Cairo, Egypt.

[4]. Debapratim Sarkar, Raghunath Ghosh , (2009)"*A Bottom-Up Approach of Line Segmentation from Handwritten Text*".

[5]. N. Chaddha, R. Sharma, A. Agrawal, and A. Gupta., (1995)"*Text Segmentation in Mixed-Mode Images*". In 28th Asilomar Conferenc e on Signals, Systems and Computers, pages 1356-1361.

[6]. Karin Sobottka, Horst Bunke and Heino Kronenberg, (1999) "Identification of Text on Colored Book and Journal Covers", Document Analysis and Recognition, 20-22.

[7]. Rama Mohan Babu, G., Srimaiyee, P. Srikrishna, A., "*Text Extraction From Hetrogenous Images Using Mathematical Morphology*" Journal Of Theoretical And Applied Information Technology© 2005 - 2010 Jatit.

[8]. Jui-Chen Wu · Jun-Wei Hsieh · Yung-Sheng Chen, (2008)"*Morphology-based text line extraction*" Machine Vision and Applications19:195–207 DOI 10.1007/s00138-007-0092-0.

[9]. Kohei Arail, Herman Tolle (2011)," Text Extraction From Tv Commercial Using Blob Extraction Method", International Journal Of Research And Reviews In Computer ScienceVol. 2, No.3

[10].Yuming Wang, Naoki Tanaka, (2008)"*Text String Extraction from Scene Image Based on Edge Feature and Morphology*", Document Analysis Systems.

[11]. Tesseract OCR Engine. http://vietocr.sourceforge.net/

[12]. Hicham EL MOUBTAHIJ, Akram HALLI, Khalid SATORI,(2014) "*Review Of Feature Extraction Techniques For Offline Handwriting Arabic Text Recognition"* International Journal of Advances in Engineering & Technology, Vol. 7, Issue 1, pp. 50-58 ISSN: 22311963

AUTHORS

D. R. Mehta received M.E. degree from Veermata Jijabai Technological Institute (VJTI), Mumbai in 1993. He is currently Associate Professor of VLSI at VJTI,Mumbai. His research interest includes Bio-Medical Image Processing and VLSI Circuits.

Amit B. Kore received the Bachelor's Degree from Sinhgad Academy of Engineering, Pune university in 2010. He is currently pursuing toward the M.Tech Degree in VJTI at Mumbai University. His current research interest include Content Based Image retrieval (CBIR), Text Character Recognition Systems and Audio Processing.