IOSR Journal of VLSI and Signal Processing (IOSR-JVSP)
Volume 3, Issue 2 (Sep. – Oct. 2013), PP 55-61
e-ISSN: 2319 – 4200, p-ISSN No. : 2319 – 4197
www.iosrjournals.org

Automatic detection and recognition of Malayalam text from natural scene images

Rosemol Emmanuel[1], Jilu George[2]
[1] *Electronics and Communication Engineering, FISAT, MG University, Kerala, India*
[2] *Electronics and Communication Engineering, FISAT, MG University, Kerala, India*

Abstract: *In this paper we describe a very simple and efficient method for the détection and recognition of the Malayalam text from colour natural scene images taken by a mobile phone camera. Malayalam text detection, skew correction of the detected text ,text segmentation and character recognition are the important steps in text understanding from natural scene images. Text understanding in natural scene image is very important for many purposes such as assistance system for the visually challenged persons and text translation in foreign countries. The experimental results show that our method can successfully extracts and recognizes the text with low complexity and therefore can be used in the mobile devices which have limited capability.*
Keywords : *Skew angle estimation, text detection, text segmentation, text recognition.*

I. Introduction

Malayalam text detection and recognition in real images taken in unconstrained environments, such as street view images, sign board images etc. remain surprisingly challenging in Computer Vision. Malayalam text detection and recognition in natural scene images has gained much attention in the last few years as it is a primary step towards fully autonomous Malayalam Optical Character Recognition (OCR) system. Images of natural scenes have numerous difficulties compared to the traditional scanned documents. These may contain diverse complex text of different sizes, styles and colours with complex backgrounds. Furthermore, such images are captured under variable lighting conditions and are often affected by the skew distortion. Text understanding in natural scene image is very important because it helps to make an assistance system for visually disabled persons and also helps to make text translator in foreign countries. Text detection, skew correction of detected text, text segmentation and character recognition are the important steps in text understanding from natural scene images. Block diagram of the proposed text detection and character recognition system is shown in Fig 1:

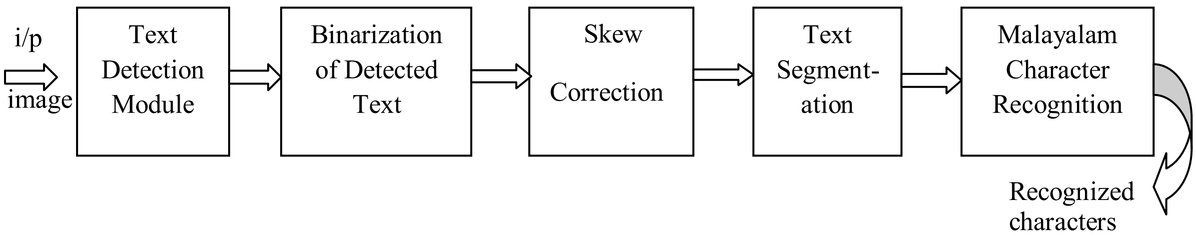


figure 1: Block diagram of the proposed system

The main Malayalam text in colour natural scene images is detected by combining an edge profile based method and some features of the Malayalam characters. Here the text area of interest is find out from the histogram of the edge image. For text binarization, the colour images are usually converted to grayscale images. The binarization techniques for grayscale images are of two types: global binarization and local binarization. Otsu method [2] which tries to find a single threshold value for whole text image, is used for the text binarization. This is one of the famous global binarization technique. The skew estimation and correction of the text image is crucial as it affects the line segmentation of the text. Here we used the bottom profile method for the skew angle estimation [3] and corrected this skew by rotating the text with this skew angle. Text segmentation is crucial as it is the primary step towards the character recognition system. In this paper we have proposed a bounded box method for segmentation of text image into lines, words and characters [4 There are two basic methods used for character recognition namely Matrix Matching and Feature Extraction. Matrix matching method works well for our application. [5]. The foreigners use mobile phone to capture the signboards of shops, traffic signboards etc. In our Malayalam text translating system, the Malayalam text in the image is extracted and translated to other languages. The experimental results show that our method is very efficient and suitable for applications in mobile devices.

II. Text Area Detection

Text understanding has gained considerable attention over the past few years, particularly, in the case of street scenes. This problem has manifested itself in various forms, namely, object detection and object recognition. Several text detection methods have been proposed based on edge profile, binarization, spatial-frequency image analysis etc. Here we presented an edge profile based method to detect the text region from the natural scene image. The input of the system is a colour image captured by a mobile phone camera (SAMSUNG GALAXY) which is shown in Fig 2

figure 2:input image

2.1 Median filtering of the gray scale image

The conversion of RGB image to gray scale image has been done because the further processing is done on a gray scale image. First action to be performed is the median filtering of the gray scale image. In image processing, it is often desirable to be able to perform some kind of noise reduction on an image. The median filter is a nonlinear digital filtering technique, often used to remove noise. Such noise reduction step improves the results of later processing (for example, edge detection on an image). The main idea behind the median filtering is to run through the image pixel by pixel, replacing each pixel with the median of neighboring pixel. The pattern of neighbors is called the "window", which slides, entry by entry, over the entire signal. If the window has an odd number of entries, then the median is simple to define: it is just the middle value after all the entries in the window are sorted numerically. For an even number of entries, there is more than one possible median. Note that this step blurs the background more than text so the edge of text will be retained. This is shown in Fig 4 and Fig 5.

5	7	10
12	11	8
1	4	15

⇒ 1, 4, 5, 7, <u>8</u>, 10,11, 12, 15 ⇒

median

5	7	10
12	<u>8</u>	8
1	4	15

figure 3: Filtering approach of median filter

figure 4:gray scale image

figure 5:median filtered image

2.2 Conversion of median filtered image to edge image

Usually the signboard text is horizontal, so the horizontal edge profile in the edge image is calculated to detect the text region. Sobel edge-emphasizing method is used to create the edge image. The Sobel operator performs a 2-D spatial gradient measurement on an image which emphasizes regions of high spatial gradient that correspond to edges. Sobel operator uses two 3 by 3 kernels, one for horizontal changes and the other for vertical changes. If X is the original image, then approximations of the derivatives - one for horizontal changes (G_X), and one for vertical(G_y), are obtained by convolving these kernels with the original image X.

$$G_X = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{pmatrix} * X \qquad (1)$$

$$G_Y = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix} * X \qquad (2)$$

Where * denotes the two dimensional convolution operation

Here G_X is defined as increasing in the "right"-direction and G_Y is defined as increasing in the "down"-direction. At each point in the image, the resulting gradient approximations can be combined to give the gradient magnitude, using:

$$G = \sqrt{(G_X^2 + G_X^2)} \qquad (3)$$

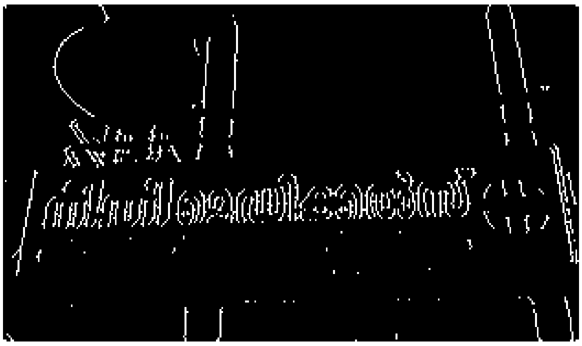
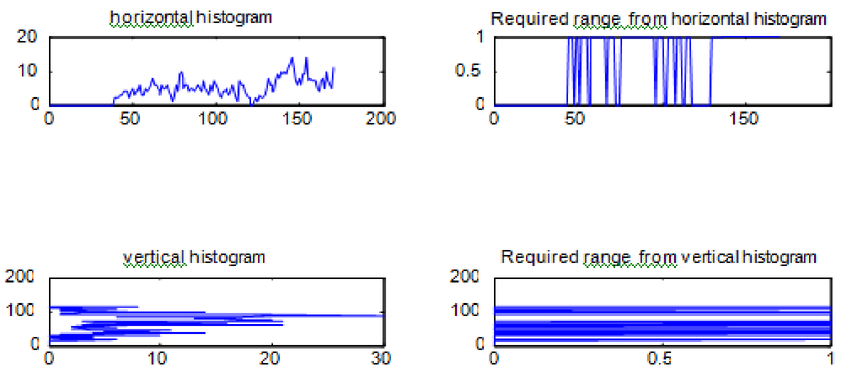


figure 6: sobel edge image

2.3 Text area detection

By using the histogram plot, the supporting idea to extract the text region of interest can be implemented. This is done by concentrating on the central area of the image, which helps to exclude the unwanted informations. This can improvised by looking into the hope of that the height of the text area of interest was almost protruding out than the other textures, which included even the minor details. The images analyzed showed an optimal result with images which were collected. The histogram of the edge image is shown in Fig 7 and the detected text region is given in the Fig 8



(a) Horizontal and vertical histogram (b) selected histogram range from (a)

figure: 7

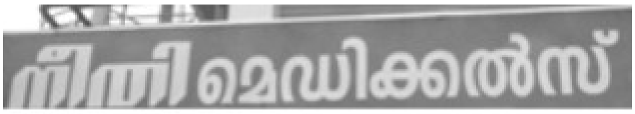


figure 8: Detected text region

III. Binarization

The goal of binarization is to convert the given input grayscale or colour image into a bi-level representation (black and white). Binarization of the image simplifies the segmentation of the text. The binarization techniques for grayscale images are of two types: global and local binarization. Global binarization techniques try to find a single threshold value for whole text image. The local binarization methods have been proposed to overcome some problems like illumination over the text image is not uniform etc. Since the size of text image is small, global thresholding method seems efficient to be used to extract text from the background. The famous global binarization method is Otsu [5]. Otsu method uses clustering based image thresholding. This algorithm assumes that the image to be binarized contains two types of pixels (foreground and background) and then found out the optimum threshold separating these two types of pixels so that their combined spread is minimum. The pixels of given image be represented in L gray levels [1,2,…L] and the total number of pixels by $N = n_1 + n_2 + \ldots + n_i$. Here the gray-level histogram is normalized and regarded as a probability distribution for simplicity. One can see that, the simple and fast Otsu method can generate good results for normal images

Figure 9: Binarized image

IV. Skew Angle Estimation And Correction

The skew angle estimation and correction of the text image is very important because it may affect the segmentation processes to be followed. A number of skew angle detection and correction methods are available for the document images. Projection profile method and Hough transform method work well for the text only documents. A common disadvantage of both these methods is that the computational complexity is proportional to the accuracy required. Here we used the bottom profile method [3] for the skew angle estimation and correction. Fig 10 shows the logic behind this method:

figure 10: Bottom profile method

In this method we assume a horizontal straight line at the bottom of the image and calculated the heights in terms of pixels from that straight line to the first white pixel found while moving upward. These heights are measured along the complete width of the text region. Among the heights, find the leftmost (*h1*), rightmost (*h2*) heights and by using these heights find the skew angle. Thus, we can find the skew angle as:

$$\text{Skew angle} = \text{inverse tan}\left(\left(\left(h2-h1\right)/d\right)\right) \quad (4)$$

Where d is the total width of the text. After finding the skew angle, rotate the image by this skew angle to obtain the skew corrected image.The skew corrected image is shown in the Fig 11:

figure 11: Skew corrected image

V. Text Segmentation

Text segmentation into lines, words and characters is a critical step towards the text recognition system. Each character extracted is the input to the Character Recognition System. Global horizontal projection method, piece-wise horizontal projection analysis, Hough transform and thinning operation are the commonly used text segmentation methods. Here we have proposed a bounded box method for segmentation of documents lines and words and characters [4].

5.1 Line Segmentation

Line segmentation is required, if the text in the natural scene image contains more than one text line. The steps for line segmentation are as follows:

(1) Count the black pixel in each row of the image.
(2) Find the rows containing no white pixel.
(3) Replace all such rows by 1
(4) Mark the Bounding Box for text lines.
(5) Copy the lines in Bounding Box and save it in separate file.

5.2 Word segmentation

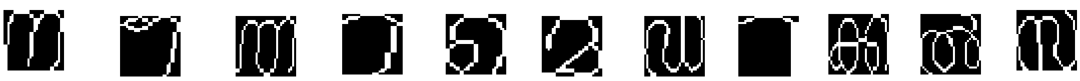
Word segmentation is required, if the text in the natural scene image containing more than one word. The steps for word segmentation are as follows:

(1) Count the black pixel in each column.
(2) Find the columns containing no white pixel.
(3) Replace all such columns by 1
(4) Mark the Bounding Box for word.
(5) Copy the words in the Bounding Box and save it in separate file.

Figure 12: word segmented image

5.3 Character segmentation

Character segmentation uses connected component labelling. A set of pixels in which each pixel is connected to all other pixels is called a connected component. Connected components labelling scans an image and groups its pixels into components based on pixel connectivity, i.e. .all pixels in a connected component share similar pixel intensity values and are in some way connected with each other. In this method, each untouched characters of a word is a connected component and unique label is assigned to each connected component in the binarized image. By this way we can extract the entire text in the image.



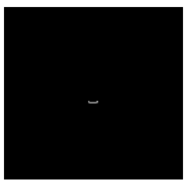


figure 13: character segmented images

VI. Malayalam Character Recognition

6.1 Malayalam language and script

Malayalam is one of the four major Dravidian languages of South India. It is syllabic in nature and alphabets are classified into vowels and consonants. Conjunct symbols are used to combine certain consonants. Malayalam language script consists of 15 vowels (Fig 11) and 36 consonants (Fig 12). Even though Malayalam script has been standardized, people still used to write in both old script and new script. In Malayalam language many characters are distinct just with a small variation in appearance

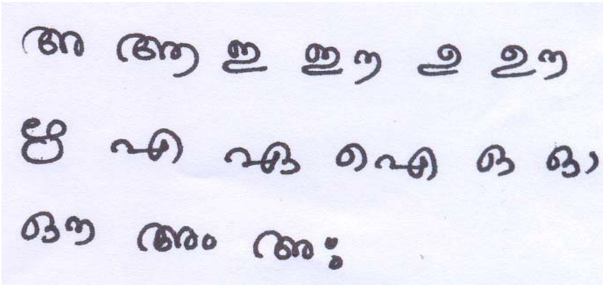


figure 14. Vowels in Malayalam

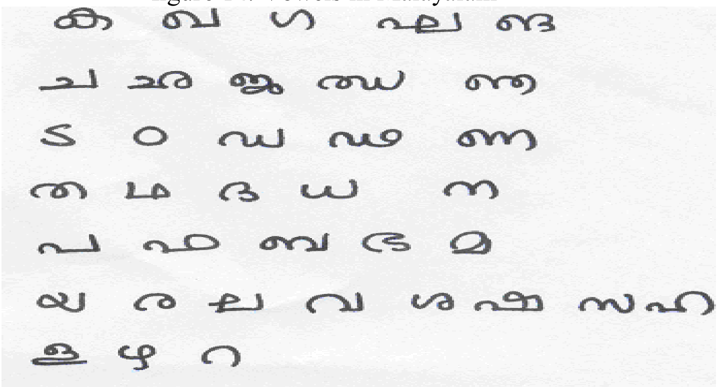


figure 15. Consonants in Malayalam

A good recognition system should be capable of recognizing characters in different fonts, styles and sizes. Modern Malayalam character recognition methodologies have enabled the recognition of complex characters and symbols .Here we use the matrix matching (template matching) approach for the recognition of Malayalam characters.

6.2 Template matching method

Here we used template matching method for the character recognition process. Here individual image pixels are used as features. Recognition is performed by comparing an input character with a set of templates (or prototypes) from each character class which is stored in a separate file. Each comparison results is a similarity measure between the input characters with a set of templates. After all templates have been compared with the observed character image, the character's identity is assigned the identity of the template which is having highest percentage of similarity. The purpose of this paper is to successfully recognize and digitize Malayalam characters from natural scene images and store them with ease.

For template matching, correlation between the main image and the template image was being used. The result of the matching process lies in range 0 to 1 which indicates 0% and 100% matching respectively. For the research implementation, we use 2-dimensional cross correlation in template matching technique. The 2-dimensional cross correlation equation is shown below:

$$\text{Out}(i,j) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} in1(m,n)\; in2(m,n) \qquad (5)$$

where in1 and in2 are the input images and out is the result of the cross correlation between in1 and in2. in1 operates as the main image and in2 as the template image. In this, template image is matched and shifted with the main image until it finds the highest matching percentage between both the images[5]. This process is repeated until all the characters in the main image is compared with template image. The template image which is having highest correlation value is selected and the process is repeated for the entire text in the image. As the number of samples for each character increases, the accuracy of the recognition system is also increases. All the recognized characters, which are labeled with corresponding character codes are converted into UNICODES. UNICODE is a standard for representing characters and integers. Unlike ASCII which uses 7 bits for each character, Unicode uses 16 bits, which means that it can represent more than 65,000 unique characters. Malayalam Unicode ranges from 0D00 to 0D7F. The Unicode for some Malayalam characters are listed below:

Some post-processing techniques are applied to the recognized text, which corrects the mistakes that occurred during the recognition stage. Linguistic rules are applied to the recognized text to reduce errors. For example, certain characters don't occur at the beginning of a word and if occur, they are remapped appropriately. Similarly, dependent vowel signs can occur only with consonants or consonant conjuncts; if found along with vowels or soft consonants, they are remapped into consonants/conjuncts similar in shape to the vowel sign. Independent vowels occur only at the beginning of a word and if found anywhere else, they are remapped appropriately. The recognized text after the post-processing is then written on a notepad, as shown in Fig 13, can be used as input to a Malayalam text to speech system or a language converter.

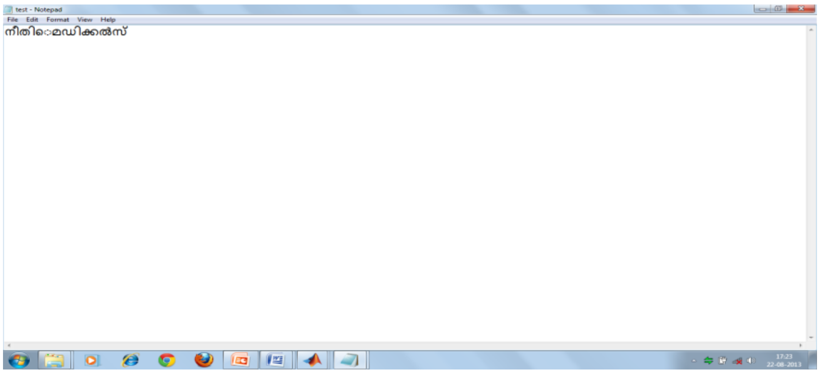


Figure 16: Recognized text written on a notepad

VII. Conclusion

In this paper, we propose an efficient method for the detection and recognition of the Malayalam text from colour natural scene images. The main text of natural scene is simply detected by combining edge profiles and characteristics of Malayalam characters specified in our application. The detected text is then skew corrected using the bottom profile method .The skew corrected image is then segmented using bounded box method. Character recognition is done using template matching method. After passing through all these stages, the recognised Malayalam characters are written on a notepad. The proposed algorithm can overcome the main challenges associated with the natural scene images like complex background, different font styles of the text, sizes of the text and orientation of the text. But it fails to overcome the problems with uneven illumination, reflection etc. The entire system can be integrated with a Malayalam text to speech system to get a text reading system for the visually challenged. If the entire system is integrated with a language converter module, it will help foreigners to get the entire text to their own language. The experimental result shows that our low-complexity approach can achieve better results in the detection and recognition of the text from natural scene images and can be used in mobile devices applications.

References

[1]. Toan Nguyen Dinh, Jonghyun Park and GueeSang Lee," Korean Text Detection and Binarization in Colour Signboards," International Conference on Advanced Language Processing and Web Information Technology, 2008

[2]. Nobuyuki Otsu, "A Threshold Selection Method from Gray-Level Histogram". IEEE Transaction on Systems, Man and Cybernetics (1979).

[3]. Ayatullah Faruk Mollah, Subhadip Basu, and Mita Nasipuri ,"Text/Graphics Separation and Skew Correction of Text Regions of Business Card Images for Mobile Devices ," Journal of computing, volume 2, issue 2, February 2010, ISSN 2151-9617

[4]. Vikas J Dongre , Vijay H Mankar "Devnagari document segmentation using histogram approach" Department of Electronics & Telecommunication, Government Polytechnic, Nagpur, India

[5]. Nadira Muda, Nik Kamariah Nik Ismail, Siti Azami Abu Bakar, Jasni Mohamad ZainFakulti Sistem Komputer & Kejuruteraan Perisian, "Optical Character Recognition By Using Template Matching " University Malaysia Pahang Karung Berkunci 12, 25000 Kuantan, Pahang