International Journal of Psychosocial Rehabilitation, Vol. 23, Issue 04, 2019
ISSN: 1475-7192

Real Time Text Detection and Shopping Assistance for the Visually Impaired

M Padma Priya, J K Periasamy, P Jayaselvi, R Aditya Siddharth

Abstract--- *The blind person is assisted while shopping so that they can buy quality products and are not misled by false claims or discounts. Here we propose a camera based system which is used to help blind person in reading text patterns printed on objects. To obtain the object from the background and extract the text pattern from that object, the system captures the image from the camera and then detects the object region. Also a recommendation system is built which helps to choose the right product based on customer review. Customer reviews can give an idea about the quality of the product, reviews are collected from online websites and a Sentiment analysis algorithm is used to analyze the review of the product recognized*

Index Terms--- *Image Processing, Opinion Mining, Optical character recognition, Sentiment Analysis*

I. INTRODUCTION

Blind people face difficulty in purchasing the right product on their own so they usually depend on their family members for shopping assistance, who often may not be able to afford quality time for them. There are many devices that can provide good access to common objects such as product packages and objects with printed text such as names of different brands. The blind person is assisted while shopping so that they can buy quality products without being misled by fake reviews and ratings. This is done to avoid the cumbersome process of exchange of product or refund of money. The text on the label is extracted and the name or brand of the product is recognized using optical character recognition. The text regions are then separated from the non-text regions using geometric properties of text. Finally the recognized characters of the text are merged to form a word and output is given in the form of audio or speech. Customer review gives an idea about the quality of the product. Sentiment analysis algorithm which uses set of rules and a dictionary and analyzes the review of the product recognized from online websites. Positive and negative expressions are identified from the review and a counter or sentiment score is initialized which is incremented or decremented based on the nature of the review and final score decides the quality of the product.

II. RELATED WORK

Text Recognition using Image Segmentation and Neural Network presented by Ammar. A.Radhi[1] uses neural networks for image extraction to enhance and optimize the process of recognizing the text in an image. Several algorithms were introduced to detect the characters from the image and an optical character recognition in order to convert any scanned document to a format that is recognizable for the computers. However, many challenges are

M Padma Priya, Department of Computer Science and Engineering, Sri Sairam Engineering college, Chennai, India.
J K Periasamy, Department of Computer Science and Engineering, Sri Sairam Engineering college, Chennai, India.
P Jayaselvi, Department of Computer Science and Engineering, Sri Sairam Engineering college, Chennai, India.
R Aditya Siddharth, Sify Technologies Ltd, Chennai, India.

Received: 09 Sept 2019 | Revised: 08 Oct 2019 | Accepted: 07 Nov 2019

accompanying the recognition from such images i.e., the text in such images is widely variable considering its layout and fonts. The neural network used in this task was trained by complete words from dictionary instead of letters.

Real Time Text Detection and Recognition on Hand Held Objects to Assist Blind People published by Samruddhi Deshpande, Revati Shriram[2] proposed an approach to identify Maximally Stable External Regions (MSER) thereby making it possible to achieve high performance using MSER and OCR for text detection and recognition

III. PROPOSED SYSTEM

The proposed system will detect the text patterns from the image .The detected text are given as speech output to blind person. Also, opinion mining is used to suggest better products, if any.

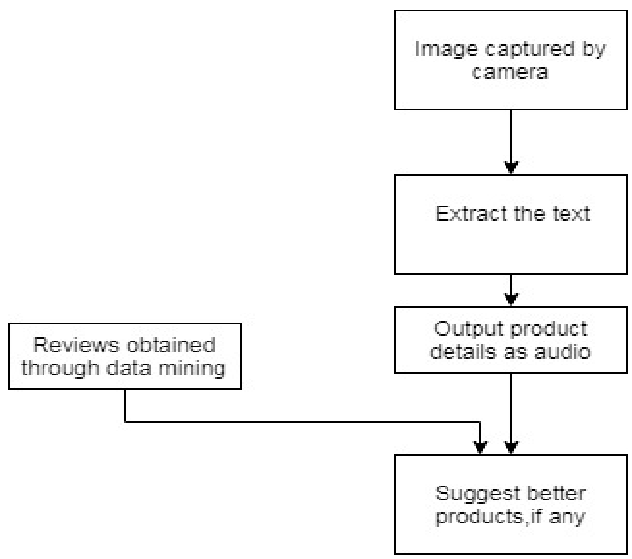

Figure 1.General Block diagram of Proposed System

IV. METHODOLOGY

A. *Optical Character Recognition*

OCR technology provides scanning systems and image systems with the ability to convert images of characters into machine letters, in characters capable of being interpreted or recognized by a computer. Thus, the images of characters in machine letters are extracted from a bitmap of the image reproduced by the scanner. The OCR process involves several aspects such as segmentation, feature extraction and classification. Image processing is an absolutely crucial field of work for those groups and industries that are working in areas such as medical diagnosis, astronomy, geophysics, environmental sciences, laboratory data analysis, industrial inspection, etc. As a first step, the image is trimmed to fit the text. After this, line by line is separated. the threshold for the binary image transformation is 0.99 (bn = im2bw (image, 0.99)). This threshold was taken so that colors with RGB values very close to 255 (maximum value) are considered as 0 in the binary image. Once each line of the image is obtained separately, we proceed to extract letter by letter from the matrix image. For this, the bwlabel function was used, which labels the connected components of the image. In other words, this function counts the continuous traces and lists them. The main operation that was used for the classification was the correlation in two dimensions. This operation gives a value of the similarity between two matrices (images).

MATLAB's mathematical and creation environment is ideal for image processing, since these images are, after all, matrices. This toolbox incorporates functions for:

- Filter design.
- Improvement and retouching of images.
- Analysis and statistics of images.
- Morphological, geometric and color operations.
- 2D transformations.

The recognized text is stored as a string. The string is produced as an output in the form of speech. For this purpose, Microsoft Speech Application Programming Interface is used. The SAPI API provides a high-level interface between an application and speech engines. SAPI implements all the low-level details needed to control and manage the real-time operations of various speech engines. The two basic types of SAPI engines are text-to-speech (TTS) systems and speech recognizers. TTS systems synthesize text strings and files into spoken audio using synthetic voices. Speech recognizers convert human spoken audio into readable text strings and files.

Applications can control text-to-speech (TTS) using the ISpVoice Component Object Model (COM) interface. Once an application has created an ISpVoice object the application only needs to call ISpVoice::Speak to generate speech output from some text data.

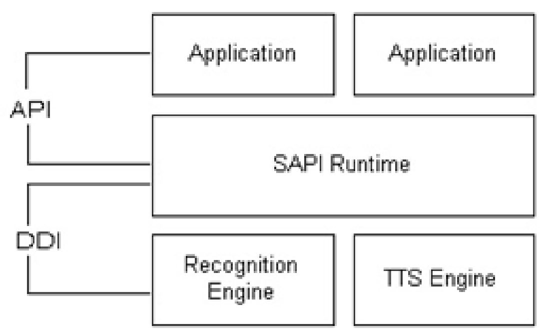

Figure 2. SAPI architecture

Special SAPI controls can also be inserted along with the input text to change real-time synthesis properties like voice, pitch, word emphasis, speaking rate and volume. This synthesis markup sapi.xsd, using standard XML format, is a simple but powerful way to customize the TTS speech, independent of the specific engine or voice currently in use.

International Journal of Psychosocial Rehabilitation, Vol. 23, Issue 04, 2019
ISSN: 1475-7192

The IspVoice::Speak method can operate either synchronously (return only when completely finished speaking) or asynchronously (return immediately and speak as a background process). When speaking asynchronously (SPF_ASYNC), real-time status information such as speaking state and current text location can polled using ISpVoice::GetStatus. Also while speaking asynchronously, new text can be spoken by either immediately interrupting the current output (SPF_PURGEBEFORESPEAK), or by automatically appending the new text to the end of the current output.

In addition to the ISpVoice interface, SAPI also provides many utility COM interfaces for the more advanced TTS applications.

B. *Sentiment Analysis*

Detailed detection of the aspects of a product can be quite challenging. Sometimes the aspect is explicit. The comment can be classified as a negative or positive expression. A structure must be chosen to analyze the reviews given by the customers. The approach used in the system is to divide the text into a list of sentences. The next step is to tokenize these sentences. Tokens are identified from each sentence. Each token is considered as a three-element tuple. The tuple consists of a word form,word lemma and a list of tags associated with the token. The word form is the word that appears in the text. The word lemma is the generalization of the word.

The text is preprocessed as in NLP. NLTK library is used for this purpose. To recognize positive and negative expressions, dictionaries are used, i.e. simple files containing expressions that will be searched in our text. To quantify the sentiment measure, positive and negative expressions are counted. But this measure ignores the fact that some expressions maybe more positive or more negative than others. A way of defining this "strength" could be using two new dictionaries: one for "incrementers" and another for "decrementers". Some expressions could be incorrectly tagged. A negation can be used to convey a positive opinion. But the negation used may lead to wrong conclusion. This is resolved by taking into account these types of polarity flips defining a dictionary of inverters. The algorithm is implemented using Python. Using this analysis, products can be compared based on their merits and demerits. A product with a sentiment score is considered better and beneficial.

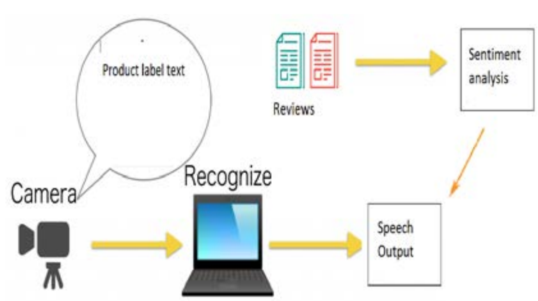

Figure 3. Proposed System Architecture

V. EXPERIMENTAL RESULTS

First of all the product is scanned and MSER regions are detected. Maximally stable Extremal regions(MSER) are used as a method of blob detection in images. This method of extracting a comprehensive number of corresponding image elements contributes to the wide-baseline matching, It performs continuous transformation of image coordinates. The MSER feature detector works well for finding text regions. It works well for text because the consistent color and high contrast of text leads to stable intensity profiles. The detect MSER Features function is used to find all the regions within the image and plot these results. Although the MSER algorithm picks out most of the text, it also detects many other stable regions in the image that are not text.

Figure 4 shows the MSER region detection after scanning the product.

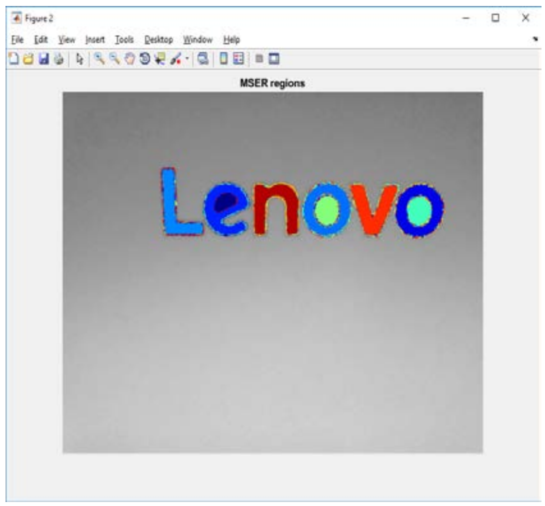

Figure 4.Detecting MSER Regions

A rule-based approach is used to remove non-text regions. For example, geometric properties of text can be used to filter out non-text regions using simple thresholds. Figure 5 shows the result of discriminating between text and non-text regions after removing Non-textual regions based on geometric properties. Geometric properties can be the aspect ratio, Euler number, Extent etc.

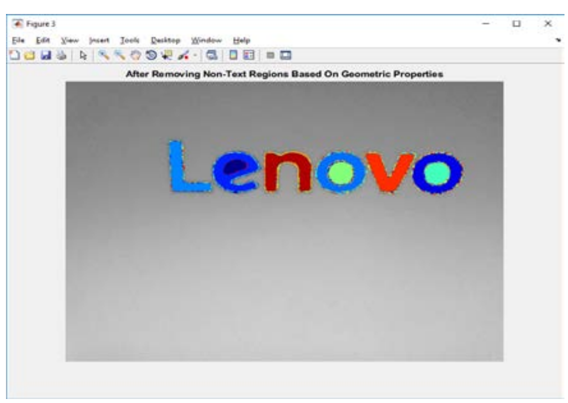

Figure 5.Removing non-textual regions

Received: 09 Sept 2019 | Revised: 08 Oct 2019 | Accepted: 07 Nov 2019

Another common metric used to discriminate between text and non-text is stroke width. Stroke width is a measure of the width of the curves and lines that make up a character.

Text regions tend to have little stroke width variation, whereas non-text regions tend to have larger variations. Figure 6 shows how the stroke width has been detected from the image.

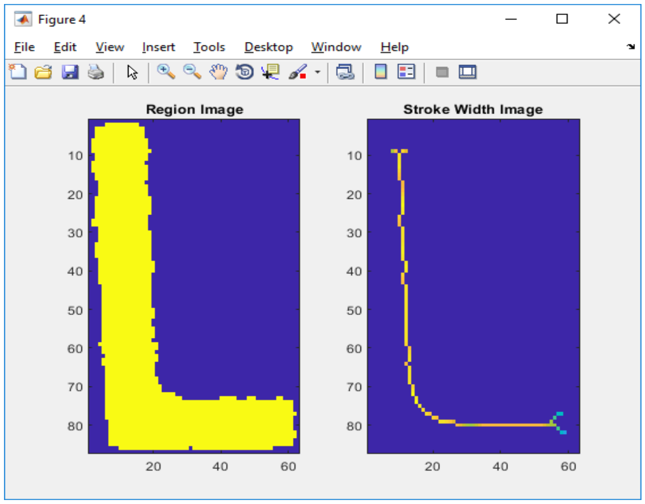

Figure 6.Stroke detection

In order to use stroke width variation to remove non-text regions using a threshold value, the variation over the entire region must be quantified into a single metric. Then, a threshold can be applied to remove the non-text regions. The individual text characters must be merged into words or text lines. This enables recognition of the actual words in an image, which carry more meaningful information than just the individual characters. Figure 7 shows the resultant image after removing Non-Text Regions based on Stroke Width variation

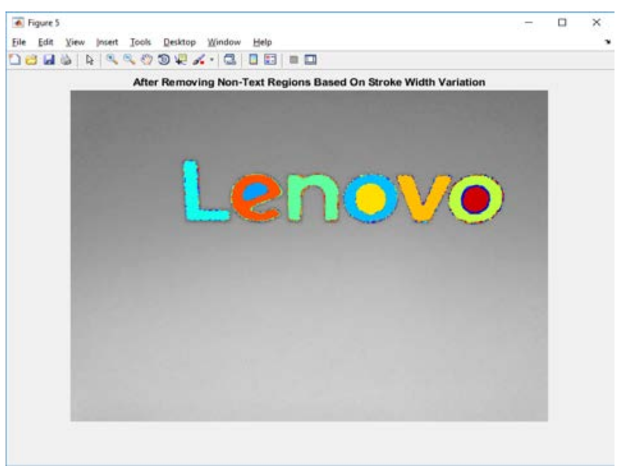

Figure7.Finding Text based on Stroke width variation

Next step is to merge individual text regions into words or text line to find neighboring text regions and then form a bounding box around those regions. To find neighboring regions, expand the bounding boxes. This makes the bounding boxes of neighboring text regions overlap such that text regions that are part of the same word or text line form a chain of overlapping bounding boxes. Now, the overlapping bounding boxes can be merged together to form a single bounding box around individual words or text lines.Figure 8 shows the result of expanding bounding boxes around the text.

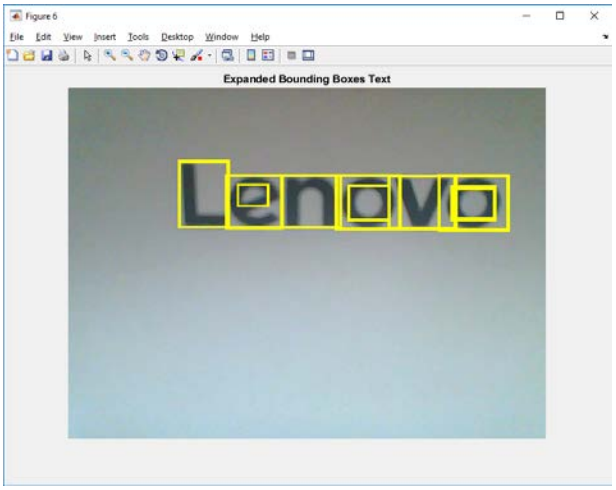

Figure8.Bounding boxes

Finally False text detections are suppressed by removing bounding boxes made up of just one text region. This removes isolated regions that are unlikely to be actual text given that text is usually found in groups (words and sentences). After detecting the text regions, the OCR function is used to recognize the text within each bounding box. Figure 9 shows the result of text detection after expanding bounding boxes.

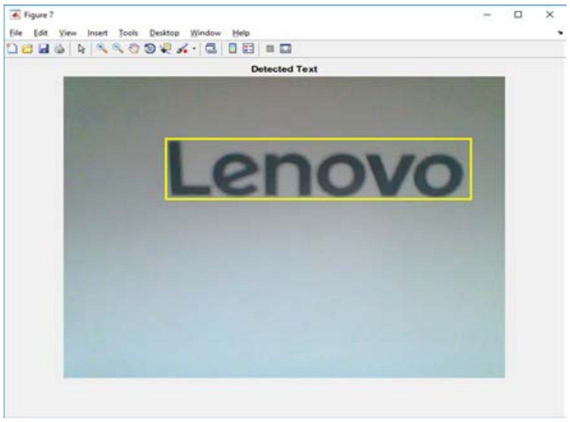

Figure9.Text Detection

VI. CONCLUSION

The proposed system can efficiently detect the product name and accurately predict its quality based on customer reviews. The system scans the label of the product and eliminates the non-textual regions based on stroke width variation. Individual characters are recognized with the help of bounding boxes. Then, the individual characters are merged to detect the word which is given as output in the form of audio. Further, sentiment analysis is applied to customer reviews where the text is split into sentences, and each sentence into tokens. Some basic extraction rules over the tagged text, in form of python functions were implemented and the sentiment measure is calculated. Based on the sentiment score, the quality of the product is decided. It has the perspectives of ubiquitous deployments and advanced improvements.

The proposed system can be further improved by improving the accuracy of the sentimental analysis algorithm using classification and machine learning algorithms that can be applied for better interpretation of reviews using natural language processing.

REFERENCES

[1] Ammar A. Radhi,“ Text Recognition using Image Segmentation and Neural Network,” International Journal of Computer Applications (0975 – 8887) Volume 170 – No.10, July 2017.

[2] Samruddhi Deshpande, Ms. Revati Shriram, “Real Time Text Detection and Recognition on Hand Held Objects to Assist Blind People”, 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT) International Institute of Information Technology (I²IT), Pune.

[3] Chucai. Yi , Y. Tian. and Areis . Arditi, ”Portable Camera-Based- Assistive-Text-and-Product Label Reading From Hand-Held Objects for Blind Persons”IEEE/ASME TRANSACTIONS ON MECHATRONICS.

[4] B. Epshtein, E. Ofek, and Y. Wexler, Detecting text in natural scenes with stroke width transform, in Proc. Comput. Vision Pattern Recognit., 2010, pp. 29632970.

[5] Wararat Songpan, “The analysis and prediction of customer review rating using opinion mining**”,** IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA),2017