CONVERSION OF IMAGE INTO TEXT TO REGIONAL TEXT AND SPEECH

BalaKrishnan K., Shivapraksh Ranga, Nageswara Gupth

Department of ISE Bengaluru

Bala.k.btech@gmail.com; shivaprakashranga@gmail.com; mnguptha@yahoo.com

Abstract— In the current scenario digital images are the raw inputs for all the digital image processing applications. These raw images are captured by the electronic devices which contains the unwanted noise or background scene. Preprocessing technique is used to completely eliminate the unwanted noise from an image. In the existing research work, conversion of image to text and speech are analyzed and results are obtained which is useful to the various image processing applications and visually challenged people. In the proposed practice, text to particular regional language text conversion is done which is useful to the Tours and travels and visually challenged people. Using this methodology, they can convert the images into their regional language like bus boards, shop boards and product etc., These regional language text are converted into speech signals using TTS.

Index Terms— Canny Edge Detection, Preprocessing, Text Detection, Text Recognition, Speech to Text.

I. INTRODUCTION

In the current scenario the digital images play a quite important role in the digital image processing applications such as restorations and enhancements, digital cinema, image transmission and coding, color processing, remote sensing, robot vision, hybrid techniques, facsimile, pattern recognition, registration techniques, multi-dimensional image processing, video processing, high resolution display, high quality color representation and super high definition image processing. Image processing enhances the raw images from the electronic gadgets in our day to today life. These raw images are used for various applications in the digital image processing research area. These raw images are taken as an input and performed with preprocessing. The raw image contains some unwanted noise or background scenes which is harmful to examine the images so that de- noising techniques such as filtering, enhancement, detection and localization are used to identify the edges. The Image segmentation and Edge detection are the two important techniques of preprocessing. The low level image engineering process is the inceptive process of image segmentation. This method is ultimately used to simplify the raw digital images into segments or pixels which is comfortable to examine and identify the effective edges in a complex image [1]. The image engineering process is classified into three categories namely low level, medium level and high level engineering. In the low level image engineering process, raw digital images are transformed into pixels. These pixels are modified into attributes in the mid-level engineering process and these attributes are altered into symbols in the high – level image engineering process. Based on the low level image engineering process, the pixels are used to identify the effective edges. Edges are defined as the crossing of two edges [2]. Edges are the important feature in edge detection techniques in numerous digital image processing applications such as object recognition, motion analysis, pattern recognition, computer- guided surgery, finger print recognition, automatic traffic controlling systems, anatomical structure and image processing. Detecting the edges from noisy images and corrupted images are difficult. In the past two decades' diverse edge detection techniques are proposed and used to recognize the effective edges. These edge detection techniques prevent the problems of false edge detection, edge localization and computational time [3]. The detection of edges contains the various algorithms such as Sobel operator, Prewitt operator, Roberts operator, Laplacian of Gaussian, Optimal edge detector (Canny algorithm). In that canny edge detection algorithm provides the optimal result with respect to complex images. In this proposed methodology canny edge deduction algorithm is used to measure the effective edges.

Text recognition is used to retrieve the text from the raw images automatically. The detection of text contains the following approaches such as text detection & localization and text recognition. Several approaches for text detection in images and videos have been proposed in the past. Based on the methods being used to localize text regions, these approaches can be categorized into three main classes such as region based methods, texture based methods and hybrid approaches. To recognize the texts from an image a texture based mechanisms are used. The image is translated into text format then theses text format is converted into particular regional language text based on the user's requirements. The conversion of text to regional language text is done by using Google API. Generally, the conversion of image to text is performed with only the particular language due to most of the information or text captured from the image will be in particular language. That information can be read or understand by the person who knows that particular language and others feel difficulty in understanding that information even though after converting image to text. Hence converting image to default text is a major issue. To overcome the issue here we deal with the solution of converting default language text to particular language text of user choice. TTS (Text to Speech) is used to convert the particular text into speech. Some people doesn't know to read the particular regional language

and it will be useful to the visually challenged people also.

The rest of this paper is organized as follows. Section II discuss the proposed methodology. Section III focuses on Experiment and Results. Section IV presents the conclusion and feature enhancement.

I. METHODOLOGY

In the past two decades the image processing research focuses on image to text conversion and text to speech conversion. The main advantage of these research helps to the visually challenged people. In the proposed conversion methodology, it will help the people who doesn't know the particular regional language in our country. India has more regional languages in the world. The main advantage of this methodology to help the people in tours and travels and visually challenged people. Generally, the image to text conversion is done only with the particular language for example the image is taken which contains the English text means it converts only into English text and it will be converted into speech using TTS. Due to overcome the disadvantages we proposed the new methodology is shown in the fig 1. In this we have following phases like

- Preprocessing
- Text Detection and Localization
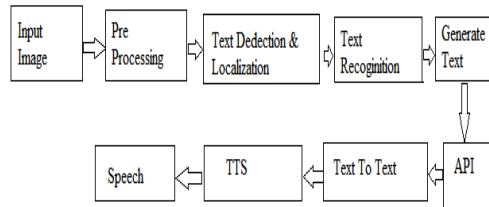- Text Recognition
- Text to Text API
- Text To Speech (TTS)

Fig 1. Methodology

Preprocessing

The preprocessing is required for all the digital image processing applications to eliminate the noise from an image. Normally the captured image contains some unwanted backgrounds or noise. Images are segmented in the form of pixels or masks. To exclude the noise, the de-noising techniques are used which is given below [4]
- Filtering
- Enhancement
- Detection
- Localization

Filtering is a technique for reforms an image. It is used to provide or improve for better performance result of an edge detector with respect to noisy images. It is used to highlight the features or remove some features in the image.

Image enhancement is the process of modify the digital images into pixels for analysis. The following operations are performed eliminate noise, sharpen an image, or brighten an image which is making it simple to identify key features of an image and used to compute the gradient magnitude with respect to pixels.

Detection process finds the exact edge point in the segmented image. Threshold value is used detect the exact edges in the pixels.

Localization is used to evaluate the resolution with respect to sub pixel. The preprocessing steps are given in the Fig 2.
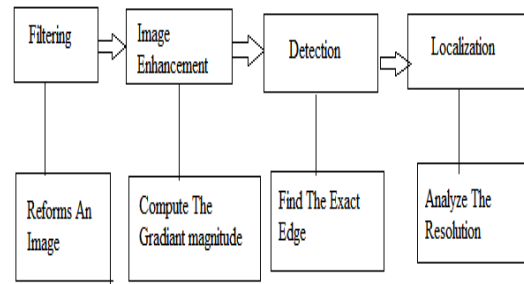

Fig 2. De noising Techniques

Edge detection

Edges are important feature in an image. Finding the effective edges with respect to noisy or corrupted images are difficult due to lighting conditions, background, luminance and geometrical features, noise volume, missing to detect existing edges and false edge. The distinct edge detection techniques are developed to expose the edges of an image. These techniques are classified with respect to pixels and sub pixels. Early edge deduction techniques are classical operators (Sobel operator, Prewitt operator, Robert's cross operator and Kirsch operators) all are used to compute the first derivative to distinguish the edges. Laplacian Gaussian operator was introduced which uses the Gaussian function to compute the second order gradients of an image. The Gaussian or optimal edge detection focuses on finding the sharp edges with fixed characteristics in all directions and sensitive to noise [5]. The classification of edge detection, purpose, merits and demerits are shown in the Table 1.

The canny edge detection algorithm is used to identify the effective edges with respect to noisy images. the edge detection steps are given below [6] and process is shown in the Fig 3.

Step1: Smooth the digital image and gradient filter.

Step 2: Calculate the gradient(M) of the particular image.

Step 3: The direction of the edge is computed using the gradient in the x and y directions.

Step4: Relate the edge direction to a direction that can be traced in an image.
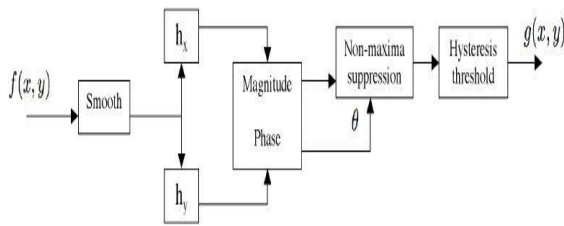
Step5: Non – maximum suppression.

Step6: Hysteresis.

Fig 3. Steps for Canny Edge detection algorithm

Technique	Purpose	Merits	Demerits
Classical	To compute the first order derivative	Simplicity Find the smooth edges	Inaccurate Average results
Laplacian of Gaussian	To compute the second order derivative	Find the exact edges Check the pixels in wide area	Few edges cannot be detected
Gaussian Filter (canny edge Detection)	To compute the optimal edge detection	Finding error rate is high	False Zero crossing

Table 1. Purpose, Merits and Demerits of edge detection techniques

Text detection and recognition

The detection and localization of the text from the images are difficult task in the digital image processing. The key challenges are identified and categorized into three types such as Diverse nature of scene text, Complexity of the background and interfacing elements. Optical Character Recognition (OCR), ABBYY Fine Reader, Microsoft Office Document Imaging(MODI) are turned up to detect and recognize the writing of text in the scanned images. The accuracy attained by the OCR is 97%. Many of the text detection methods are developed in the past two decades such as boundary, color, texture and character.

- Region Based approaches
- Texture based approaches
- Hybrid approaches
- Character based approaches

In the Region based approaches uses the basic similarity criteria like color, size, width and edges in the pixel which is not applicable to the complex images.
In the Texture based approaches uses the texture image properties to find the text from the images. The wavelet, FET, Spatial variance and globes filters are used to detect the textural properties of a text area in an image.
In the Texture based detection is able to detect even if the image is deformed
but computationally complex when the rate of edge is high. In the Hybrid approaches, uses both region based and texture based approaches in the single images which found a location.

Color feature based text detection is able to detect the inclined and deformed licensed plates but the finding of RGB is limited in condition.
In Character based text detection is robust in rotation but time consuming production with complex images [7].

Text to speech conversion

A text to speech converter is used to convert the normal language text into speech[TTS]. It can be achieved in both the hardware and software. This application is useful in day to day applications such as Banking, Railways and Visually challenged people [9]. Text to Speech synthesizer is used to convert the text data into speech signals. The conversion of input text into linguistic representation is called as text to phonetic or grapheme conversion which is shown in the Fig 4.
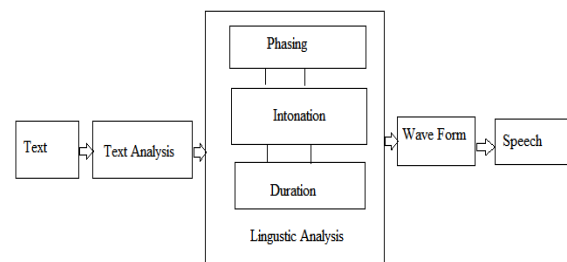

Fig 4. TTS Architecture

In this, three types of acoustic synthesis are used to convert the text to speech signals are [8]
- Concatenative synthesis
- Format synthesis
- Articulatory synthesis

In this concatenative synthesis contains the prerecorded human voice for all the words. TTS gives the natural voice translation is the advantage but the drawback is larger database is needed to store the words.
Format synthesis uses the concept of artificial and robotics.it doesn't have any prerecorded voice.
Articulatory synthesis model which is mainly used in mathematical models.

II. EXPERIMENT & RESULTS

This paper presents the image conversion to text to regional language text and speech. They are tested with verity of natural images and their corresponding edges are identified using preprocessing techniques and the particular text is converted into malayam regional language text and that particular text is converted in the form of speech signals using TTS which is shown in the Fig 6. Finally, 95 % of accuracy is obtained in both regional level text as well as speech signals.

Fig 6 Conversion of Text to Text and Speech

III. CONCLUSION & FUTURE ENHANCEMENT

In this paper, Image is converted into text and that text is converted into particular regional language text through the proposed system architecture. In this canny edge detection algorithm is used to get the optimal result with respect to complex images and TTS is used to convert the speech signals. The proposed algorithm is evaluated and the accuracy obtained by the result is 95%. Finally, it is recommended for the future work to test the more complex images and get the 100% accuracy in it and mobile applications may be created for effective usage of proposed methodology which is useful to the tours and travels and visually challenged people. Travelling from one country to another country the journey will be fine without depend the local regional languages.

REFERENCES

[1] Beant K and Anil G,"Mathematical Morphological Edge Detection for Remote Sensing Images" IEEE Transcations, pp.324-327, 2011J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Payal S K , Phade G M and Dhulekal P A ," Survey paper on edge and corner detection algorithms", International Journal & Maganzine of Engineering, Technoilogy, Management and reserch, Volume 2, issue 5, pp.344-351,2015.

[3] Ireyuwa E I," Comparision of Edge Detection Techniques in image processiong Techniques", International Journal of Information technology & Electrical Engineering" Volume 2, pp.25-29, 2013.

[4] Kanika W and Gajendran E, "Automated Recognition of Text in images: A survey", International Jornal of Computer Applications,Volume 127,pp.15-19,2015.

[5] Lakshmi S and Sankaranarayanan," A study of Edge Detection Techniques for Segmentation Computing Approaches", IJCA Special Issue on "Computer Aided Soft Computing Techniques for Imaging and Biomedical Applications", pp.35-41, 2010.

[6] Senthilkumar N and Rajesh R,"Edge detection techniques for image segmentation – A survey of softcomputing approaches", International Journal of Recent Trends in Engineering ", volume 1, Issue 2,2009.

[7] Mingixu L and Shuai C," A new Prediction method for edge detection based on human visual feature", IEEE 24th Chinese Control and Detection Conference, pp.1465-1468,2012.

[8] Poonam S S,S A Patil and P M Jadhav," Review of text to speech conversion methods",International Journal of Industrial Electronics and electrical Engineering,Volume 2,pp.29-35,2014.

[9] Pijus K,"Diphone database for Lithunian Text to Speech Synthesis",Iformatica,Volume   16,pp.192-202,2015.