International Journal of Engineering Sciences & Emerging Technologies, Nov. 2015.
ISSN: 22316604 Volume 8, Issue 3, pp: 80-87 ©IJESET

AN ADVANCED APPROACH ON OPTICAL CHARACTER RECOGNITION AND SPEECH GENERATION

Ashima Sindhu Mohanty, Subhrajit Pradhan, Akshya Ku Sahoo
Assistant professor, G I E T, Gunupur, Odisha, India
,

ABSTRACT

The goal of Optical Character Recognition (OCR) is to classify optical patterns corresponding to alphanumeric or other characters. OCR involves several steps including segmentation, feature extraction, and classification. Each of these steps is a field unto itself, and is described briefly here in the context of a Matlab implementation of OCR. By using this project, system can be designed to help the visually handicapped people for solving real life problems.

KEYWORDS—*MSER, Canny Edge, OCR, Connected component labelling, TTS (text to speech) System.*

I. INTRODUCTION

The paper is on optical character recognition (OCR) and speech generation based on Matlab and computer vision. Computer vision is a field that includes methods for acquiring, processing, analyzing, and understanding images and, in general, high-dimensional data from the real world in order to produce numerical or symbolic information, e.g., in the forms of decisions. OCR is the mechanical or electronic conversion of images of typewritten or printed text into machine encoded text. It is widely used as a form of data entry from printed paper data records, whether passport documents, invoices, bank statements, computerized receipts, business cards, mail, printouts of static data, or any suitable documentation. Speech Synthesis is the artificial production of human speech. A computer system used for this purpose is called a speech computer or speech synthesizer, and can be implemented in software or hardware products. A text-to-speech (TTS) system converts normal language text into speech.

II. COMPUTER VISION

A. Methods:

Computer vision is a field that includes methods for acquiring, processing, analyzing, and understanding images and, in general, high-dimensional data from the real world in order to produce numerical or symbolic information, e.g., in the forms of decisions. A theme in the development of this field has been to duplicate the abilities of human vision by electronically perceiving and understanding an image. This image understanding can be seen as the disentangling of symbolic information from image data using models constructed with the aid of geometry, physics, statistics, and learning theory. Computer vision has also been described as the enterprise of automating and integrating a wide range of processes and representations for vision perception.
As a scientific discipline, computer vision is concerned with the theory behind artificial systems that extract information from images. The image data can take many forms, such as video sequences,

views from multiple cameras, or multidimensional data from a medical scanner. As a technological discipline, computer vision seeks to apply its theories and models to the construction of computer vision systems. Subdomains of computer vision include scene reconstruction, event detection, video tracking, object recognition, object pose estimation, learning, indexing, motion estimation, and image restoration [3].

B. Applications:

Applications range from tasks such as industrial machine vision systems which, say, inspect bottles speeding by on a production line, to research into artificial intelligence and computers or robots those can comprehend the world around them. The computer vision and machine vision fields have significant overlap. Computer vision covers the core technology of automated image analysis which is used in many fields. Machine vision usually refers to a process of combining automated image analysis with other methods and technologies to provide automated inspection and robot guidance in industrial applications.

In many computer vision applications, the computers are pre-programmed to solve a particular task, but methods based on learning are now becoming increasingly common. Examples of applications of computer vision include systems for:

1. Controlling processes, *e.g.*, an industrial robot;
2. Navigation, *e.g.*, by an autonomous vehicle or mobile robot;
3. Detecting events, *e.g.*, for visual surveillance or people counting;
4. Organizing information, *e.g.*, for indexing databases of images and image sequences.

III. MAXIMALLY STABLE EXTREMAL REGIONS

In computer vision, maximally stable external regions (MSER) are used as a method of blob detection in images. This technique was proposed by Matas et al. to find correspondences between image elements from two images with different viewpoints. This method of extracting a comprehensive number of corresponding image elements contributes to the wide-baseline matching, and it has led to better stereo matching and object recognition algorithms.

In the field of computer vision, blob detection refers to mathematical methods that are aimed at detecting regions in a digital image that differ in properties, such as brightness or color, compared to areas surrounding those regions. Informally, a blob is a region of a digital image in which some properties are constant or vary within a prescribed range of values [1, 2, 10].

IV. CANNY EDGE DETECTOR

The Canny edge detector is an edge detection operator that uses a multistage algorithm to detect a wide range of edges in images [5, 6]. It was developed by John F. Canny in 1986. Canny also produced a computational theory of edge detection explaining why the technique works.

The Process of Canny edge detection algorithm can be broken down to 5 different steps which are as follows:

1. Apply Gaussian filter to smooth the image in order to remove the noise
2. Find the intensity gradients of the image
3. Apply non maximum suppression to get rid of spurious response to edge detection
4. Apply double threshold to determine potential edges
5. Track edge by hysteresis: Finalize the detection of edges by suppressing all the other edges that are weak and not connected to strong edges.

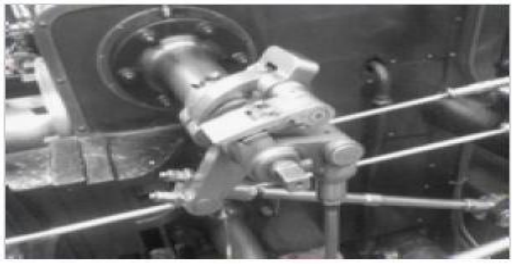


Fig. 1 Original image of the steam engine

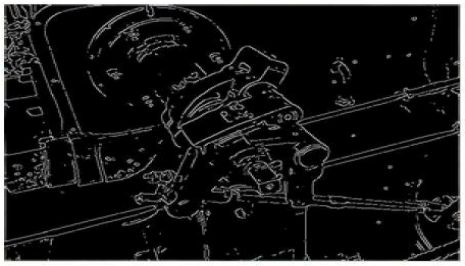


Fig. 2 Canny edge detector applied to image of a steam engine

Fig.1 and Fig.2 the application of Canny edge detection technique to a steam engine.

V. CONNECTED COMPONENT LABELLING

This (alternatively connected-component analysis, blob extraction, region labeling, blob discovery, or region extraction) is an algorithmic application of graph theory, where subsets of connected components are uniquely labeled based on a given heuristic. Connected component labeling is not to be confused with segmentation.

Connected-component labeling is used in computer vision to detect connected regions in binary digital images, although color images and data with higher dimensionality can also be processed. When integrated into an image recognition system or human-computer interaction interface, connected component labeling can operate on a variety of information. Blob extraction is generally performed on the resulting binary image from a thresholding step. Blobs may be counted, filtered, and tracked.

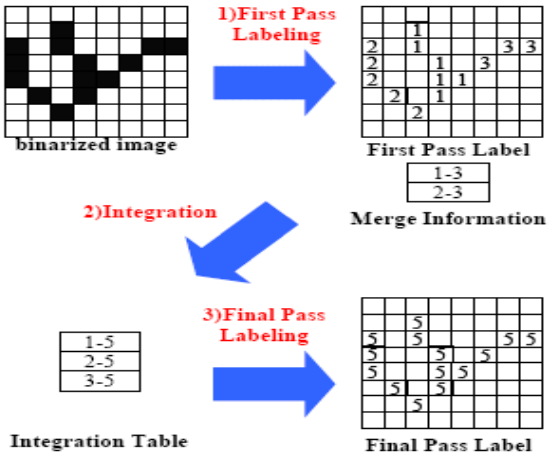


Fig. 3 Process of connected component labeling

VI. OPTICAL CHARACTER RECOGNITION

Optical character recognition (OCR) is the mechanical or electronic conversion of images of typewritten or printed text into machine encoded text. It is widely used as a form of data entry from printed paper data records, whether passport documents, invoices, bank statements, computerized receipts, business cards, mail, printouts of static data, or any suitable documentation. It is a common method of digitizing printed texts so that it can be electronically edited, searched, stored more compactly, displayed online, and used in machine processes such as machine translation, text to speech, key data and text mining [7, 8].

Early versions needed to be trained with images of each character, and worked on one font at a time. Advanced systems that have a high degree of recognition accuracy for most fonts are now common. Some systems are capable of reproducing formatted output that closely approximates the original page including images, columns, and other non-textual components.

A. Applications:

OCR engines have been developed into many kinds of object oriented OCR applications, such as receipt OCR, invoice OCR, and check OCR, legal billing document OCR.
They can be used for:

- Data entry for business documents, e.g. check, passport, invoice, bank statement and receipt
- Automatic number plate recognition
- Automatic insurance documents key information extraction
- Extracting business card information into a contact list
- More quickly make textual versions of printed documents, e.g. book scanning for Project Gutenberg
- Converting handwriting in real time to control a computer (pen computing)
- Defeating CAPTCHA anti-bot systems, though these are specifically designed to prevent OCR
- Assistive technology for blind users.

B. Block diagram of OCR:

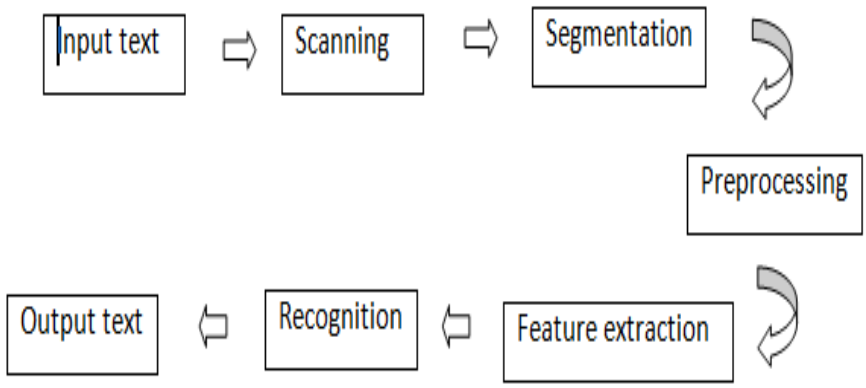


Fig. 4 represents the block diagram of OCR and the different processes involved in it.

- **Scanning**

Scanning makes original document as digital image. Generally, original documents are made up of the black colored text print on the white colored background. Scanning comes with thresholding which makes the digital image as gray scale image. Thresholding is the process which converts multi level image into bi-level image i.e. black and white image. Fixed threshold level is defined in thresholding. If the gray levels are below the threshold level, identified as black, whereas if gray level is above the threshold level, identified as white [4]

- **Segmentation**

The process of locating regions of printed or handwritten text is segmentation. When segmentation is applied to text, it isolates characters or words. The mostly occurred problem in segmentation is: it causes confusion between text and graphics in case of joined and split characters. If document is dark photocopy or if it scanned at low threshold, joints in characters will occur. And splits in characters will occur if document is light photocopy or scanned at high threshold [4].

- **Preprocessing**

Some noise may occur during scanning process. This results in poor recognition of characters. This usually occurred problem is overcome by preprocessing. It consists of smoothing and normalization. In smoothing, certain rules are applied to the contents of image with the help of filling and thinking techniques. Normalization is responsible to handle uniform size, slant and rotation of characters.

- **Feature Extraction**

It extracts the features of symbols. In this, symbols are characterized and unimportant attributes are left out. The feature extraction technique does not match concrete character patterns, but rather makes note of abstract features present in a character such as intersections, open spaces, lines, etc.

- **Recognition**

OCR system recognizes characters. It identifies characters in foreground pixels, called as blobs, and then it finds lines. Word by word recognition of characters is done throughout the lines. Recognition extracts text from images of documents.

C. Types:

- Optical character recognition (OCR) – targets typewritten text, one glyph or character at a time.
- Optical word recognition – targets typewritten text, one word at a time (for languages that use a space as a word divider). (Usually just called "OCR".)
- Intelligent character recognition (ICR) – also targets handwritten print script or cursive text one glyph or character at a time, usually involving machine learning.
- Intelligent word recognition (IWR) – also targets handwritten print script or cursive text, one word at a time. This is especially useful for languages where glyphs are not separated in cursive script.

D. Technique:

- De-skew- If the document was not aligned properly when scanned, it may need to be tilted a few degrees clockwise or counter clockwise in order to make lines of text perfectly horizontal or vertical.
- Despeckle – It removes positive and negative spots, smoothes edges.
- Binarization –It converts an image from colour or greyscale to black-and-white (called a "binary image" because there are two colours). In some cases, this is necessary for the character recognition algorithm; In other cases, the algorithm performs better on the original image and so this step is skipped.
- Line removal –It cleans up non-glyph boxes and lines.
- Layout analysis or "zoning" – It identifies columns, paragraphs, captions, etc. as distinct blocks. Especially important in multicolumn layouts and tables.
- Line and word detection – It establishes baseline for word and character shapes, separates words if necessary.
- Script recognition – In multilingual documents, the script may change at the level of the words and hence, identification of the script is necessary, before the right OCR can be invoked to handle the specific script.
- Character isolation or "segmentation" – For per character OCR, multiple characters that are connected due to image artefacts must be separated; Single characters that are broken into multiple pieces due to artefacts must be connected..
- Alignment – Segmentation of fixed pitch fonts is accomplished relatively simply by aligning the image to a uniform grid based on where vertical grid lines will least often intersect black areas. For proportional fonts, more sophisticated techniques are needed because whitespace between letters can sometimes be greater.

VII. SPEECH GENERATION

Speech Synthesis is the artificial production of human speech. A computer system used for this purpose is called a speech computer or speech synthesizer, and can be implemented in software or hardware products.

A text-to-speech (TTS) system converts normal language text into speech; Other systems render symbolic linguistic representations like phonetic transcriptions into speech.

A. Speech synthesizer:

Synthesized speech can be created by concatenating pieces of recorded speech that are stored in a database. Systems differ in the size of the stored speech units; a system that stores phones or diaphones provides the largest output range, but may lack clarity. For specific usage domains, the storage of entire words or sentences allows for high quality output. Alternatively, a synthesizer can

incorporate a model of the vocal tract and other human voice characteristics to create a completely "synthetic" voice output.

The quality of a speech synthesizer is judged by its similarity to the human voice and by its ability to be understood clearly. An intelligible text-to-speech program allows people with visual impairments or reading disabilities to listen to written works on a home computer.

B. Text-to-speech(TTS) system:

A text-to-speech system (or "engine") is composed of two parts: a frontend and a backend.
The frontend has two major tasks. First, it converts raw text containing symbols like numbers and abbreviations into the equivalent of written out words. This process is often called text normalization, pre-processing, or tokenization. The frontend then assigns phonetic transcriptions to each word, and divides and marks the text into prosodic units, like phrases, clauses, and sentences. The process of assigning phonetic transcriptions to words is called texttophoneme or graphemetophoneme conversion.
The backend is often referred to as the synthesizer, then converts the symbolic linguistic representation into sound. In certain systems, this part includes the computation of the target prosody (pitch contour, phoneme durations), which is then imposed on the output speech [9].

VIII. TECHNIQUES USED

A. Matlab programming language:

For optical character recognition and speech generation many methods can be adapted. But for our convenience we are using MATLAB (matrix laboratory) which is a multi-paradigm numerical computing environment and fourth-generation programming language. Developed by MathWorks, we are using computer vision toolbox and image processing toolbox. We are giving input image either directly from the system on which MATLAB is installed or from live camera or from a streaming video.
For detecting and recognizing text MATLAB has provided OCR function. Images provided by the webcam or static image are having certain noise which needs to be removed.

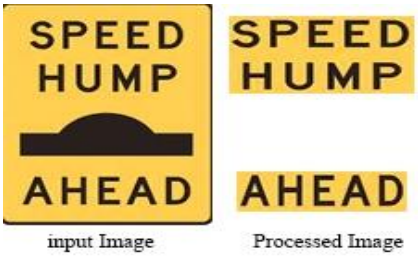
B. Process:

First MSER region is detected using detects MSER features function. since text characters usually have consistent colour, we begin by finding regions of similar intensities in the image using the MSER region detector. Canny edges work well with MSER features for text detection. So we are using edge function for detecting edges so that boundary lines of text portion can be detected.
Then by using connected component analysis, characters can be filtered. MATLAB gives output in terms of text using ocrtxt. Text function. In order to convert the written text in command window into speech a TTS (text to speech) function is required. Although it is not a predefined function in MATLAB, users of MATLAB created a function which when added into the path of MATLAB can convert written part into speech. So, finally after optical character recognition speech is produced.

IX. RESULTS

First only text region is filtered using OCR.

Example 1.



input Image Processed Image

Output at the command window of Matlab:-

SPEED
HUMP
AHEAD

TTS (Text to speech) function converts the above text part into speech.

Example 2.

After that TTS will convert the output into speech.

Input Image

Processed Image

Output at the command window of Matlab:-

BE ALERT
DONT GET HURT

X. CONCLUSION

This paper is an effort to suggest an approach for image to speech conversion using optical character recognition and text to speech technology. By this approach we can read text from a document or web page and can generate synthesized speech through a computer's speakers or phone's speaker. The developed software in MATLAB has set all policies corresponding to each and every alphabet, its pronunciation methodology, the way it is used in grammar and dictionary.

People who are totally blind or with poor vision can use this approach for reading the documents and books. Hence the application developed is user friendly, cost effective, time saving and applicable in the real time.

XI. FUTURE SCOPE

Our next works with OCR Application will include the improvement of the results by the use of advance processing techniques to detect the noise and to correct bad-recognized words. OCR application will also display the signatures and the other symbols as it is in the document. It will also update its features including the translation of one language to another. So that it will be helpful for people from other countries who can't understand the local language.

REFERENCES

[1]. Chen, Huizhong, et al. "Robust Text Detection in Natural Images with Edge-Enhanced Maximally Stable Extremal Regions." Image Processing (ICIP), 2011 18th IEEE International Conference on. IEEE, 2011.
[2]. Ray Smith. Hybrid Page Layout Analysis via Tab-Stop Detection. Proceedings of the 10th international conference on document analysis and recognition. 2009.
[3]. Reinhard Klette (2014). Concise Computer Vision. Springer.
[4]. Heuristic-Based OCR Post-Correction for smart phone Applications the university of North Carolina at chapel hill department of computer science honors thesis Author: Wing-Soon Wilson Lian 2009.
[5]. Canny, J., A Computational Approach to Edge Detection, IEEE Trans. Pattern Analysis and Machine Intelligence, 8(6):679–698, 1986.
[6]. R. Deriche, Using Canny's criteria to derive a recursively implemented optimal edge detector, Int. J. Computer Vision, Vol. 1, pp. 167–187, April 1987.

[7]. John Resig (2009-01-23). "John Resig – OCR and Neural Nets in JavaScript". Ejohn.org. Retrieved 2013-06-16.
[8]. Jump up Tappert, C. C.; Suen, C. Y.; Wakahara, T. (1990). "The state of the art in online handwriting recognition". IEEE Transactions on Pattern Analysis and Machine Intelligence.
[9]. Sproat, Richard W. (1997). Multilingual Text-to-Speech Synthesis: The Bell Labs Approach.
[10]. J. Matas, O. Chum, M. Urban, and T. Pajdla. "Robust wide baseline stereo from maximally stable extremal regions." Proc. of British Machine Vision Conference, pages 384-396, 2002.

AUTHORS

Ashima Sindhu Mohanty has completed her bachelor degree in Applied Electronics & Instrumentation Engineering and Master degree in Electronics & Instrumentation Engineering from Biju Pattanaik University of Technology respectively. She is persuing her Ph.D at SUIIT(Sambalpur University), Sambalpur, Odisha. She has 4 years of teaching experience and presently working as Assistant professor in Department of Electronics & Instrumentation Engineering, GIET, Gunupur. Her interest field includes application of fiber optics in losses control and biomedical instrumentation and application of controllers in tuning process.

Subhrajit Pradhan has completed his bachelor in Electronics & Communication Engg. and master in tech. in Electronics & Communication Engg. from Biju Pattanaik university of technology respectively. He is perusing his Ph.D at Berhampur University, Berhampur. He has 9years of teaching experience & presently he is working as assistant professor in department of Electronics & Communication in GIET, Gunupur, Odisha, India. His interest filed of research is application of neural network for signal integrity analysis of high speed digital integrated circuits.

Akshya Kumar Sahoo has completed his bachelor degree in Electrical and Electronics Engineering and Master degree in Electronics & Instrumentation Engineering from Biju Pattanaik University of Technology respectively. He is persuing his Ph.D at Berhampur University, Berhampur, Odisha. He has 8 years of teaching experience and presently working as Assistant professor in Department of Electrical Engineering, GIET, Gunupur. His interest field includes Control Systems and Renewable energy systems.