Education for the Blind – An Application of Computer Vision

M. Sowndharya, Rejoy John Benjamin

Panimalar Engineering College,

Bangalore Trunk Road, Varadharajapuram, Nasarethpettai, Ponamallee, Chennai - 600 123

sowndharya.mahesh@gmail.com, rejoy.benji@gmail.com

***Abstract* — Image recognition – renowned as computer vision – is the expertise of adopting and examining images, to produce quantitative and qualitative information. It is an approach of importing what is naturally reinforced in the human eyes and the brain into a computer's processor. Text-to-speech is a technology that converts text into speech. In this paper, a new interactive learning tool which appends these two technologies is called forth. This tool, when reading a text using TTS software, senses an image, employs image recognition technique to materialize the image into text. This content is then converted to speech. The system undergoes various phases such as pre-processing, feature extraction, object recognition, edge detection, image segmentation and text-to-speech (TTS) conversion. An algorithm to convert graphical material to speech is incorporated along. The proposal is a modest endeavour to help in the education of the unsighted.**

***Keywords* — Text to speech, Image recognition, Pattern recognition, Graph Digitizer.**

I. INTRODUCTION

285 million people are estimated to be visually impaired worldwide: 39 million are blind and 246 have low vision.

Blind people, all over the world face troubles in accessing printed material. There is a growing awareness that, the education that they receive is failing them. The urge to create new devices and tools that serve as non-visual alternatives has become important. In this way, quality education can be provided to them so as to prepare them to compete in the demanding high tech economic society of the 21st century.

Until this day, a number of assistive technological tools (Screen Reading software, special talking, and Braille devices) have been developed to help the visually challenged on a regular basis, including helping them use the internet and its resources efficiently.

The Text-to-Speech is a technology that converts text to speech. Text-to-speech is becoming more of a common feature on web sites, providing users the option to read the text on their own or have it read to them. For people who have difficulty decoding text, text-to-speech software can be used to support reading of digital text including but not exclusive to Word documents, email, accessible PDFs, and information on the Internet.

Though Screen reader technology like TTS, convert natural language text to speech, they fail to interpret the images between the texts.

An Image recognition technology can be used along with the TTS in this case. Image recognition, also known as computer vision, is the method of acquiring, analyzing, and understanding mages to produce numerical information. It is a method of importing what nature built with the human eyes and brain into a computer's processor.

In Image recognition, focus is made on a particular field called Pattern recognition that deal with recognizing anything ranging from everyday objects to scenes describing a person's action.

The real problem arises when the printed text contains graphical material within it. Various methods have evolved that aim to solve the problem, ranging from those based on technology to the construction of models, which the learner relying on tactile representation, can employ.

II. LITERARY SURVEY

Yi-Ren Yeh, Chun-Hao Huang, and Yu-Chiang Frank Wang present a approach for solving cross domain pattern recognition problem by domain adaption techique where processing and recognition of data and features are done on different domains.

Fan-Chieh Cheng, Shih-Chia Huang, and Shanq-Jang Ruan proposed the mechanism of removing background archetype form video sequence to expose foreground and objects from any applications such as traffic security, human machine interaction, object recognition and so on.

Iasonas Kokkinos and Petros Maragos construct the synergy between image segmentation and object recognition using Expectation-Maximization (EM) algorithm. An iterative operation is applied to perform these two tasks. These two tasks are performed iteratively, simultaneously segmenting an image and reconstructing it in terms of objects. Objects are modelled using Active Appearance Model (AAM) as they capture both shape and appearance variation.

III. PROPOSED WORK

In this paper, an ideal method is proposed to convey graphical material to visually impaired students. A Digitizer is a software that converts the bar-graph, scatter plots and scanned lines into a series of numbers. These numbers and scanned at regular intervals to produce a textual interpretation of the graph.

When the text to voice convertor is reading the text, it employs the algorithm to convert the graphical representation to text, on noticing a graph.

Here, to identify graphs, a graph digitizer software like Plot Digitizer or Engauge is used.

III.1. SYSTEM ARCHITECTURE

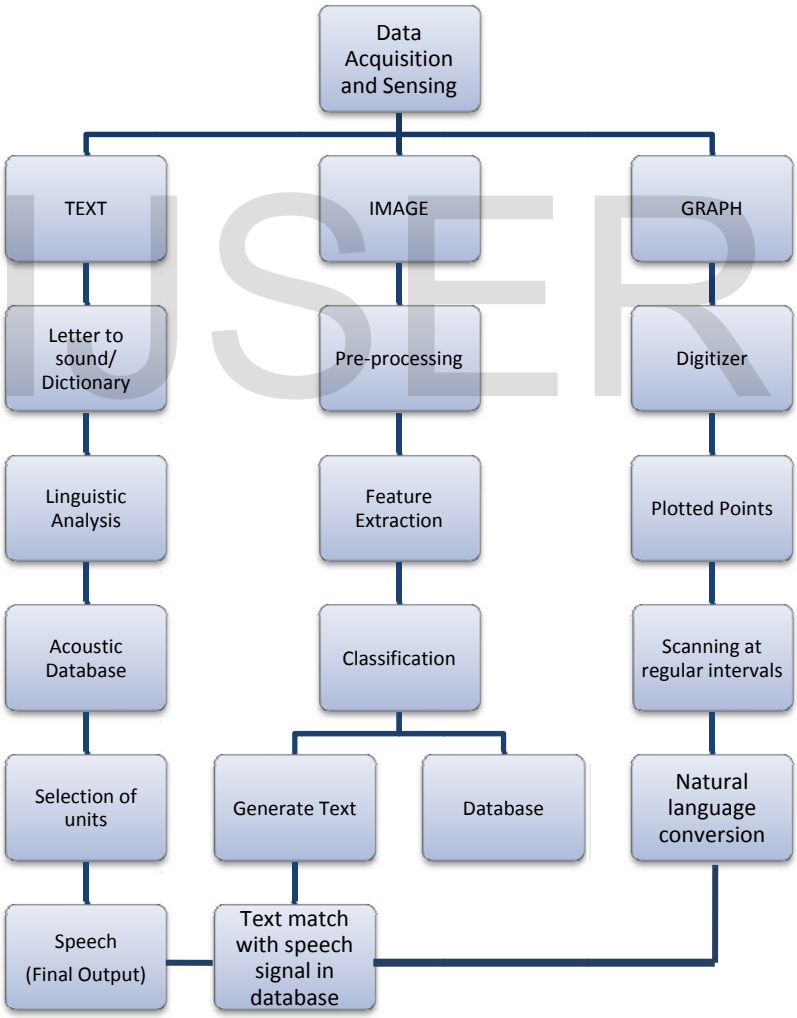

Fig. 1 System Architecture of Image to text and text to speech

IV. TEXT – TO – SPEECH

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.

Initial stage of text to speech conversion is to record the voice. In order to reproduce the natural sound of each language, a speaker records a long series of texts of all nature. Those texts contain every possible sound in the chosen language. These recordings are then sliced and organised into an acoustic database. During the database creation the recorded speech is segmented into the following: diphones, syllables, morphemes, words, phrases and sentences.

Traditionally a text-to-speech system is assembled through two parts: a front-end and a back-end. The front-end involves two major functions. First, it converts raw script comprising symbols like numbers and abbreviations into the matching counterpart of written words. This process is often called *text normalization*, *pre-processing*, or *tokenization*. The front-end then allocates phonetic transcriptions to each word, and divides and marks the text into prosodic units, like phrases, clauses, and sentences. This method of allotting phonetic transcriptions to words is called *text-to-phoneme* or *grapheme-to-phoneme* conversion.

The output is the phonetic transcriptions and prosody information conjointly make up the symbolic linguistic. The representation of an utterance that uses symbols to represent linguistic information such as information about phonetics, phonology, morphology, syntax, or semantics is referred to as linguistic representation. The back-end is the *synthesizer*. It converts the symbolic linguistic rendition 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.

V. IMAGE RECOGNITION

Image recognition is the expertise of adopting and examining images, to produce quantitative and qualitative information. It is an approach of importing what is naturally reinforced in the human eyes and the brain in a computer's processor

Image processing and image analysis tend to focus on 2D images, how to transform one image to another, e.g., by pixel-wise operations such as contrast enhancement, local operations such as edge extraction or noise removal, or geometrical transformations such as rotating the image. This characterization implies that image processing/analysis neither require assumptions nor produce interpretations about the image content.

The phases in this stage are pre-processing, feature extraction, object recognition, edge detection and image segmentation.

The inceptive step in image recognition is Noise reduction. Noise reduction is the process of removing noise from a signal. All recording devices, both analogue and digital, have traits which make them susceptible to noise. A set of connected pixels that forms a boundary between two disjoint regions is known as an edge. The task of segmenting an image into regions of discontinuity is done using edge detection. Edges usually occur on the boundary of two different partitions in an image. Edge detection helps to clearly identify the changes in region of an image where gray scale and texture change in the regions of an image. Canny algorithm concentrates predominantly on three main objectives of low error rate, minimize distance between real edge and detected edge and minimum response i.e. one detector response per edge to detect the edges in an image.

Image segmentation is a further high-ranking aspect necessarily entailed to apportion an image into regions or categories which then aids in identifying identify precisely the object in an image. Segmentation functions on the properties shown by the pixels in an image, every pixel which belongs to same category has similar gray scale value whereas pixels of different categories have dissimilar values. Segmentation is often one of the pivotal steps in examining the images because

additional overhead of moving to each new pixel of an image while working with object in an image. The task of implementing other stages becomes more facile once Image segmentation is done. While considering a fully automatic conversion algorithm, the success of image segmentation is partial and sometimes requires manual intervention.

VI. GRAPH PROCESSING

Digitizing or digitization is the representation of an object, image, sound, document or a signal (usually an analogue signal) by a discrete set of its points or samples. The result is called *digital representation* or, more specifically, a *digital image*, for the object, and *digital form*, for the signal. For a document, the term means to trace the document image or capture the "corners" where the lines end or change direction.

McQuail identifies the process of digitization has immense significance to the computing ideals as it "allows information of all kinds in all formats to be carried with the same efficiency and also intermingled".

VI.1 GRAPH DIGITIZER

Report data based on graphs must be digitized to obtain plotted points. A plot digitizer digitizes scanned images of a plot by clicking mouse on the data point. Plot digitizer can digitize with both linear, logarithmic and scaled drawings and orthographic notes. E.g. Plot digitizer. It is a digitizer software written in Java language.

It also semi-automatically digitizes lines of a plot. An image vectorization program called auto-trace is implemented here.

In the proposed algorithm, the digitized plots and the elements of X and Y axes are given as input. The algorithm performs operations like sorting and analyzing constant regular intervals and produces natural language interpretation of the entire graph.

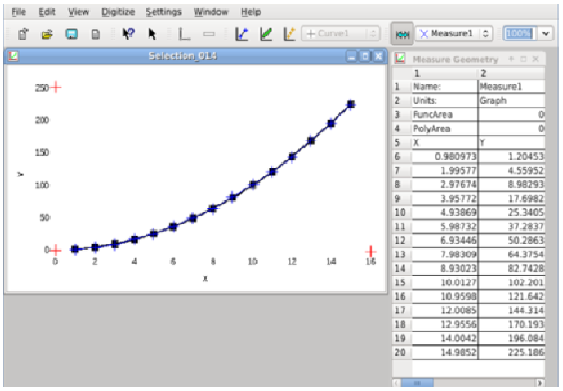

Fig. 2 Output of digitization

VI.2. ALGORITHM

From the points that are obtained from the plot digitizer, we propose an algorithm to convert the points into speech.

Input: Plotted Points.

Process: Filter points at regular intervals.

Find the elements in the X and Y axis.

Combine these points and convert to natural language.

e.g. 2010 average rainfall 200 cm.

Then employ the text to speech conversion software.

VII. CONCLUSION

Thus a number of ways to transcend an image to sound through text have been analyzed. More importance has been given to image recognition. Future work is aimed at developing more efficient algorithms on the textual interpretation of graphs.

REFERENCES

[1] Benjamin Z. Yao, Xiong Yang, Liang Lin, Mun Wai Lee and Song-Chun Zhu, “I2T: Image Parsing to Text Description” IEEE Conference on Image Processing, 2008 .
[2] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta and R. Jain, “Content-based image retrieval at the end of the early years,” IEEE Transactions PAMI, vol 22, no. 12, 2000.
[3] Y. Rui, T. S. Huang, and S. F. Chang, “Image retrieval: Current techniques, promising directions, and open issues,” Journal of Visual Communication and Image Representation, vol. 10,1999.

[4] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," ACM Computing Surveys, vol. 40, no. 2, pp. 1-60, Apr. 2008.

[5] Yi-Ren Yeh, Chun-Hao Huang, and Yu-Chiang Frank Wang, "Heterogeneous Domain Adaptation and Classification by Exploiting the Correlation Subspace," IEEE Transactions on Image Processing, vol. 23, no. 5, May 2014.

[6] Fan-Chieh Cheng, Shih-Chia Huang and Shanq-Jang, "Illumination-Sensitive Background Modeling Approach for Accurate Moving Object Detection," IEEE Trans. On Broadcasting, vol. 57, no. 4, Dec 2011.

[7] Breen A.P., "The future role of text to speech synthesis in automated services".

[8] Tanprasert, C.; Koanantakool,T., "Thai OCR: a neural network application"

[9] Iasonas Kokkinos and Petros Maragos, "Synergy between Object Recognition and Image Segmentation using the Expectation-Maximization Algorithm", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 8, Aug. 2009.

[10] Lucas S.M., „"High performance OCR with syntactic neural networks".

[11] S.V. Rice, F.R. Jenkins, T.A. Nartker, The Fourth Annual Test of OCR Accuracy, Technical Report 95-03, Information Science Research Institute, University of Nevada, Las Vegas, July 1995.

[12] R. Smith, "A Simple and Efficient Skew Detection Algorithm via Text Row Accumulation", Proc. of the 3rd Int. Conf. on Document Analysis and Recognition (Vol. 2), IEEE 1995, pp. 1145-1148.

[13]J. Canny, "A Computational Approach to Edge Detection," Reading in Computer Visions: Issues, Problems, Principles and Paradigms, pp. 184-203.

[14] P. Felzenszwalb and D. Huttenlocher, "Pictoral Structures for Object Recognition," International Journal of Computer Vision, vol. 61, no. 1, pp. 55-79, 2005.