Journal of Innovative Image Processing (JIIP) (2019)
Vol.01/ No. 01
Pages: 31-38
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2019.1.004

A SMART IMAGE PROCESSING ALGORITHM FOR TEXT RECOGNITION, INFORMATION EXTRACTION AND VOCALIZATION FOR THE VISUALLY CHALLENGED

Dr. Samuel Manoharan,
Professor,
Department of Electronics,
Bharathiyar College of Engineering and Technology,
India.
Email: jsamuel@bcetedu.in

Abstract:

Keywords: Image Processing, LattePanda Alpha, text to speech, vocalization, OCR

1. INTRODUCTION

Over 2.2 billion people suffer from blindness and visual impairment worldwide. Of this, over 1 billion people have issues that is yet to be addressed or could be prevented at an earlier stage. Most of the visually impaired people are above the age of 65. Low income countries has majority of the visually impaired population [1]. Several techniques are implemented to assist the blind and visually challenged. Pen computing software achieve 80-90% accuracy based on the handwriting. Research is being conducted to improve the accuracy and reduce the error per page in such computation.

The conversion of handwritten, printed or typed characters into machine-encoded text can be done by means of Optical character recognition (OCR). The output of this system can be manipulated by a computer. For image recognition and data entry with inputs from data records and printed documents such as business cards, resumes, passports, sales invoices, and bank statements are widely used. A text document is generated as an output by recognizing the characters in a picture or scanned document with the help of a program.

The visually impaired people are provided access to text by means of portable devices that perform video-based text acquisition [6]. Text identification from images has several challenges including system integration issues, text warping, image stabilization, and low resolution sensors [8]. There are numerous research that are directed in this domain. The navigability and usability of the system can be enhanced with the application of several cutting-edge technologies. Image processing is also used to analyse and vocalize the underlying text during mouse hovering and appraise the webpage layout structure with the help of background music and similar features. These technologies can be used for screen reader enhancement.

A system for image to speech and text conversion for the visually challenged has been developed in [9]. For detecting objects from images, Canny edge detection algorithm is used. Based on the shape, texture, size and colour of the object, it is recognized. M. Arun et al [10] use computer vision for pattern and character recognition. Digital recognition of characters from images is performed by means of OCR. A unique segmentation methodology is used for efficient and faster implementation of image processing. Selective Speech Synthesis is used for generation of speech sounds naturally based on syllables, diaphones and phonemes.

K. Ragavi et al [11] developed a portable text vocalizer that allowed audio conversion of scanned text information. In this system, the image is scanned with the help of a handheld scanner and it is sent to an android phone by means of Bluetooth. The android phone extracts the textual content from the obtained image and converts it into speech signals. An entire page with text can be scanned with a page scanner. Gerard Chollet et al [12] exploits Automatic Language Independent Speech Processing (ALISP) under video processing, image processing, text, speech and audio domains.

2. EXISTING LITERATURE

Assistive devices are categorized into Vocational training and assessment based psychological tests, low vision, daily living, vocational, mobility and educational devices. A. Karthikeyan et al [3] identified the characters in textual data by means of OpenCV. The system is divided into three stages namely –image capture and text extraction, conversion of text to speech after filtering, and conversion of speech to text. Raspberry-pi kit is used for this application.

There are several challenges in implementing text to speech conversion algorithm. Normalization of text is a complicated process. Text comprises of abbreviations, numbers, heteronyms, and so on that require to be expanded into phonetic representations. Based on the context, several similar spellings in English are pronounced in different ways. Conversion of grapheme-to-phoneme or text-to-phoneme are the basic approaches that regulates the pronunciation of a word depending on its spelling for the purpose of speech synthesis. The unavailability of a of collectively established objective assessment criteria leads to difficulty in reliable assessment of speech synthesis systems. Diverse speech data is used by different organizations.

Pravin A. Dhulekar et al [5] designed and implemented a symbol matching and character extraction based text to speech conversion system for recognition of street sign board. Such systems are already available for English street boards and signs. The system is developed for Gujarati, Marathi, Hindi and other Indian languages for multiple symbol and character identification. Ajay Roy et al [7] made use of Tesseract OCR system for the purpose of extraction of text from the scanned images. This text data is converted to speech by means of e-Speak tool for assisting the visually challenged people. The e-speak tool is used for text to speech conversion. It also assists the visually impaired person to identify products extracting the textual data from the image and performing speech conversion on the data. For this purpose, Raspberry pi is used as it provides adequate battery backup and portability.

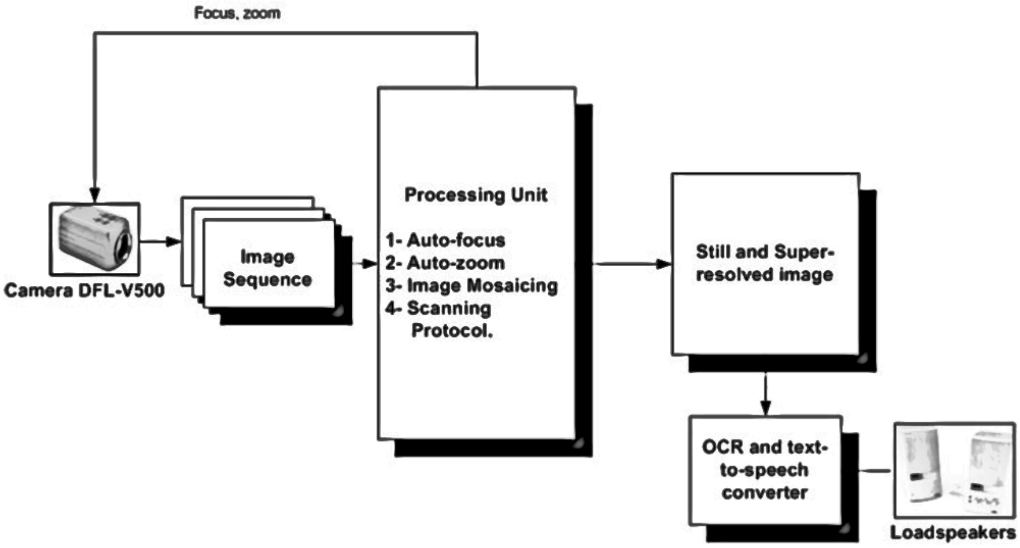


Figure 1: Video interface based text vocalizer system [6]

K.Kalaivani et al [13] introduces an efficient and innovative system for text to speech conversion that is cost efficient and can analyse both textual images and documents for the purpose of reading. The system works with MATLAB R2011b applying Text to Speech Synthesiser (TTS) and OCR concepts. Jiss Kuruvilla et al [14] studied the applications and significance of image processing in computer vision. The fundamentals of computer vision, segmentation techniques and its applications were discussed in the context of image processing. Domínguez M. Victor et al [15] used text summarization algorithm that extracts textual information from documents and perform summarization of speech-to-text. The authors investigated six algorithms for this purpose namely KLSum, SumBasic, LSA, LexRank, TextRank, and Luhn that were evaluated using OQIDSum and DUC2001 datasets.

Neha Joshi et al [16] performed text mining in a more detailed manner. The authors recognized and summarized textual information contained in an image based on the number of lines. It improves the efficiency of data and is time efficient. It works fast on huge data volumes which is time consuming and exhaustive to be performed manually. Summarization and extraction of text images is a requirement in recent days. The efficiency of the system is optimized for improved performance.

3. PROPOSED WORK

In this system, we perform digitization of the textual information with the help of scanner or camera. It is also possible to use a handheld scanner for immediate processing of information. The digitized information is processed with the help of LattePanda Alpha system on board [17] which is a tiny device that supports Linux and Windows OS. It is packed with several notable and advanced features. It offers compatibility with arduino as well as windows 10. The system disdains Edison and ARM processing powers. An Intel Celeron N4100 CPU consisting of a quad-core processor is used. Instead of windows 10 core, a complete version of windows 10 is used here. Though it is comparatively expensive with the other system on boards available in the market, it offers multiple features that serves best for this work. It also provides 2x 50-pin GPIO connectors, eMMC, microSD, SATA 3.0, PCIe x2, for use in Arduino- ATmega32U4, and so on.

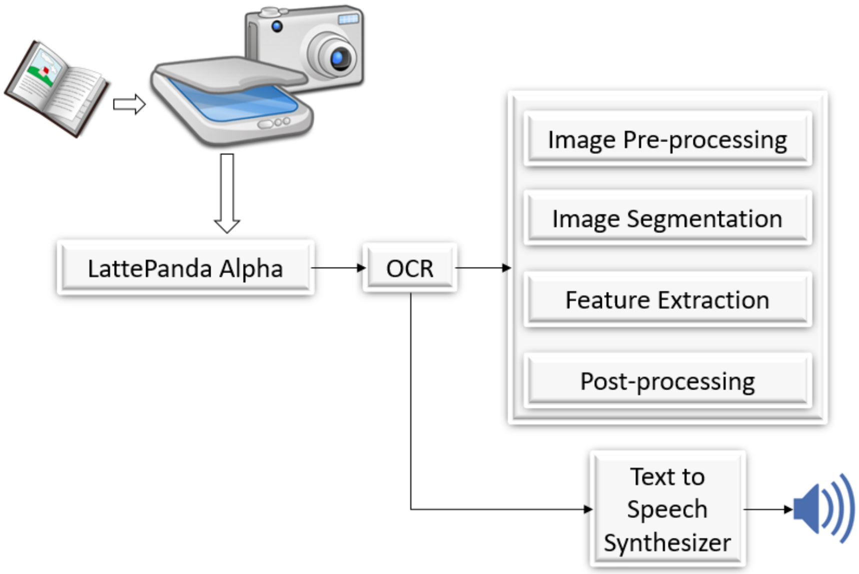


Figure 2: Architecture of the proposed vocalization system

The gathered information and digitized content is transferred to the OCR [18] for extraction of textual information from the image that can be a photo, scanned document or a video frame. This conversion happens only after the complete set of files are received by the system. The OCR uses grayscale version of the images for further analysis. Based on the intensity of the areas in the image, character identification is done. The lighter regions are considered as the background and the darker areas as text. This is further processed to identify and categorize the alphanumeric values.

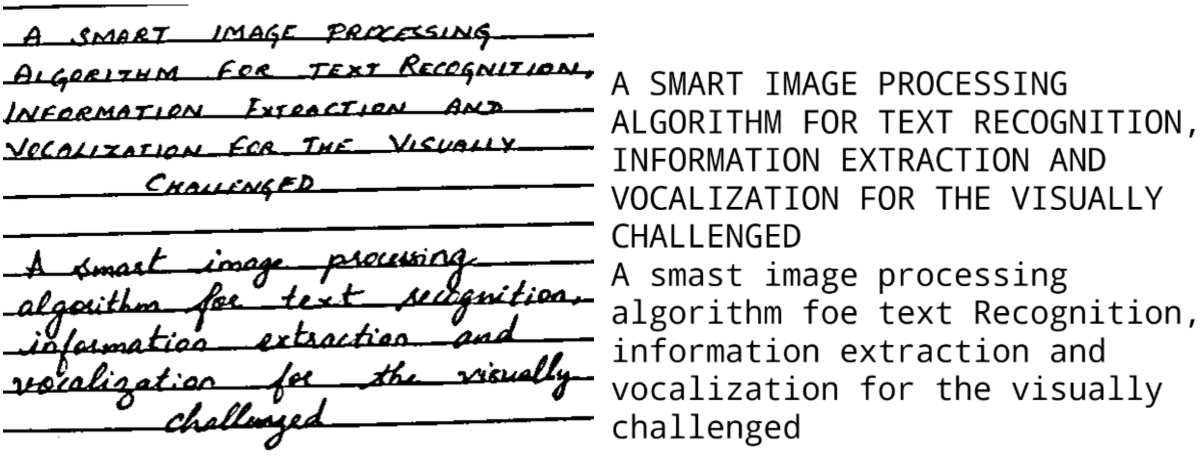
OCR pre-processing techniques include image de-skewing, de-speckling, binarization, line removal, layout analysis, line and word detection [19], script recognition, character segmentation or isolation, scaling and aspect ratio normalization. Despite the multiple techniques available, OCR operation is performed on one block of text or one word at a time. The feature detection and pattern detection algorithms are used for identification of characters. Feature detection is informed of the features of numbers or letters individually for the recognition of these characters in the document. The number of curves, crossed lines or angled lines and such features can be used for comparison of characters. For pattern recognition, sample formats and fonts of textual data are fed to the system for the purpose of recognition of characters by comparison with the scanned document.

Computer converts the identified character into its corresponding ASCII value [20]. In order to improve the accuracy of the system, handling of the complex layouts, proofreading and correction of basic errors is to be done and the document must be saved for further use. The speech synthesizer converts the final content into audio output to the visually impaired user. This system produces human speech artificially. It can be implemented in either hardware or software form. The phonetic transcription [21] and linguistic symbols representation can be converted into sound signals. An entirely synthetic model can be used for creation of voice output by replicating human voice characteristics and the vocal tract model.

The people with reading disabilities or visual impairment can thereby listen to the written data by means of a computer or mobile phone using this intelligent system. The hard disk space economy and high speed conversion is provided by using the software system. Area optimization is provided with the help of the system on board module. The output is played in the form of audio files of WAV format [22]. The data can be stored and reused if required.

4. RESULT

The system is implemented for a set of sample text images and the error rate is reduced on several iterations of the algorithm to improve the optimality of the system. The sample input and outputs of the textual data is as shown in figure 3. In case of identification of handwritten textual content, based on the readability of the content, a slight error of conversion is visible. The LattePanda Alpha system provides improved performance despite the high cost factor.



A SMART IMAGE PROCESSING
ALGORITHM FOR TEXT RECOGNITION,
INFORMATION EXTRACTION AND
VOCALIZATION FOR THE VISUALLY
CHALLENGED

A smart image processing
algorithm for text recognition,
information extraction and
vocalization for the visually
challenged

(a)

A SMART IMAGE PROCESSING
ALGORITHM FOR TEXT RECOGNITION,
INFORMATION EXTRACTION AND
VOCALIZATION FOR THE VISUALLY
CHALLENGED
A smast image processing
algorithm foe text Recognition,
information extraction and
vocalization for the visually
challenged

(b)

Figure 3: (a) Input textual image and (b) Extracted information in document format

Based on phonetics and other pre-set instructions, the textual data is converted into audio output by means of a speech synthesizer. The system provides an accuracy of 97% in identification, processing and conversion of the image input to textual and audio outputs. A time delay of 4.7 seconds is consumed for the entire process to be completed for the provided sample output.

5. CONCLUSION AND FUTURE SCOPE

The applications of speech synthesis is of significant importance and has widespread applications. Such systems help overcoming several environmental barriers for the disabled people. People with reading difficulties due to visual impairment, dyslexia, pre-literate or illiterates and so on can be benefited with this system. Communication aids with voice outputs are also employed frequently among the general public. In this system, we have used the LattePanda Alpha system for processing textual data and converting it to voice signals.

Future work is focused on improving the accuracy of the system using machine learning algorithm for processing the text information and reducing the delay time required for the processing of the information. Conversion of text from lesser learnt or minority based languages is also a major scope of research. Huge amount of memory is essential for processing the video and audio content. Reducing the size of the system while keeping the quality of the output intact is also a major challenge in implementing the system.

References

[1] Joshi, AV Kumar, T. Prabhu Madhan, and S. Raj Mohan. "Automated electronic pen aiding visually impaired in reading, visualizing and understanding textual contents." In 2011 IEEE INTERNATIONAL CONFERENCE ON ELECTRO/INFORMATION TECHNOLOGY, pp. 1-6. IEEE, 2011.

[2] Shanmugam, K., and B. Vanathi. "Hardcopy Text Recognition and Vocalization for Visually Impaired and Illiterates in Bilingual Language." In Computational Intelligence and Sustainable Systems, pp. 151-163. Springer, Cham, 2019.

[3] Karthikeyan, A., U. Kripanya, M. Manish, S. Nivetha, H. Prabanjan, and K. Ramkumar. "Cartable Camera Based Assistive Text Recognition for Visually Impaired."

[4] Monticelli, Cíntia, Regina De Oliveira Heidrich, Ronaldo Rodrigues, Ewerton Cappelatti, Rodrigo Goulart, Ricardo Oliveira, and Eduardo Velho. "Text Vocalizing Desktop Scanner for Visually Impaired People." In International Conference on Human-Computer Interaction, pp. 62-67. Springer, Cham, 2018.

[5] Dhulekar, Pravin A., Niharika Prajapatr, Tejal A. Tribhuvan, and Karishma S. Godse. "Automatic voice generation system after street board identification for visually impaired." In 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC), pp. 91-96. IEEE, 2016.

[6] Zandifar, Ali, and Antoine Chahine. "A video based interface to textual information for the visually impaired." In Proceedings of the 4th IEEE international Conference on Multimodal interfaces, p. 325. IEEE Computer Society, 2002.

[7] Rajesh, M., Bindhu K. Rajan, Ajay Roy, K. Almaria Thomas, Ancy Thomas, T. Bincy Tharakan, and C. Dinesh. "Text recognition and face detection aid for visually impaired person using Raspberry PI." In 2017 International Conference on Circuit, Power and Computing Technologies (ICCPCT), pp. 1-5. IEEE, 2017.

[8] Verma, Prabhat, Raghuraj Singh, and Avinash Kumar Singh. "A framework for the next generation screen readers for visually impaired." International Journal of Computer Applications (2012).

[9] Patil, Mrunmayee, and Ramesh Kagalkar. "An Automatic Approach for Translating Simple Images into Text Descriptions and Speech for Visually Impaired People." International Journal of Computer Applications 975 (2015): 8887.

[10] Arun, M., S. S. Salvadiswar, and J. Sibidharan. "Design and Implementation of Text To Speech Conversion for Visually Impaired Using ‘i’Novel Algorithm." Journal on Today's Ideas-Tomorrow's Technologies 2, no. 1 (2014).

[11] Ragavi, K., Priyanka Radja, and S. Chithra. "Portable text to speech converter for the visually impaired." In Proceedings of the International Conference on Soft Computing Systems, pp. 751-758. Springer, New Delhi, 2016.

[12] Chollet, Gérard, Kevin McTait, and Dijana Petrovska-Delacrétaz. "Data driven approaches to speech and language processing." In International School on Neural Networks, Initiated by IIASS and EMFCSC, pp. 164-198. Springer, Berlin, Heidelberg, 2004.

[13] Kalaivani, K., R. Praveena, V. Anjalipriya, and R. Srimeena. "Real time implementation of image recognition and text to speech conversion." Int. J. Adv. Res. Technol 2 (2014): 171-175.

[14] Kuruvilla, Jiss, Dhanya Sukumaran, Anjali Sankar, and Siji P. Joy. "A review on image processing and image segmentation." In 2016 international conference on data mining and advanced computing (SAPIENCE), pp. 198-203. IEEE, 2016.

[15] Victor, Domínguez M., Fidalgo F. Eduardo, Rubel Biswas, Enrique Alegre, and Laura Fernández-Robles. "Application of Extractive Text Summarization Algorithms to Speech-to-Text Media." In International Conference on Hybrid Artificial Intelligence Systems, pp. 540-550. Springer, Cham, 2019.

[16] Joshi, Neha. "Text Image Extraction and Summarization." Asian Journal For Convergence In Technology (AJCT) (2019).

[17] Chu, Yung-Long, Hung-En Hsieh, Wen-Hsiung Lin, Hui-Ju Chen, and Chien-Hsing Chou. "Chinese FingerReader: a wearable device to explore Chinese printed text." In ACM SIGGRAPH 2017 Posters, p. 54. ACM, 2017.

[18] Singh, Raghuraj, C. S. Yadav, Prabhat Verma, and Vibhash Yadav. "Optical character recognition (OCR) for printed devnagari script using artificial neural network." International Journal of Computer Science & Communication 1, no. 1 (2010): 91-95.

[19] Kissos, Ido, and Nachum Dershowitz. "OCR error correction using character correction and feature-based word classification." In 2016 12th IAPR Workshop on Document Analysis Systems (DAS), pp. 198-203. IEEE, 2016.

[20] Luther, Willis J., Loren A. Wood, Thomas S. Tullis, and James A. Fontana. "Method and apparatus for extracting text from a structured data file and converting the extracted text to speech." U.S. Patent 5,715,370, issued February 3, 1998.

[21] Deligne, Sabine, Francois Yvon, and Frédéric Bimbot. "Variable-length sequence matching for phonetic transcription using joint multigrams." In Fourth European Conference on Speech Communication and Technology. 1995.

[22] Whibley, Simon, Michael Day, Peter May, and Maureen Pennock. WAV Format Preservation Assessment. Technical Report. British Library. http://wiki. dpconline. org/images/4/46/WAV Assessment v1. 0. pdf, 2016.