Int J Pharm Bio Sci 2014 July ; 5 (3) : (B) 913 - 921

Review Article **Biotechnology**

International Journal of Pharma and Bio Sciences **ISSN 0975-6299**

CONVERSION OF DETECTED TEXT AND CAPTION FROM AN IMAGE TO SPEECH

R.MADHAN*[1] AND MR.S.P.CHOKKALINGAM[2]

**[1]Under Graduate Student, Department of Information Technology, Saveetha School of Engineering, Saveetha University, Chennai,*
[2]Associate Professor, Saveetha School of Engineering, Saveetha University, Chennai,

ABSTRACT

Image retrieval is a process of browsing and searching an image from large database with a collection of digital images. Images can be retrieved with the help of caption, description of the image and keyword. In the conventional image retrieval system corner detection method has been used. This corner detection method is not user friendly and also it takes more time for retrieving images when the database have a large number of images in it. In this project work we have improved the performance of the image retrieving system by applying concepts of neural networks and also combining the corner and edge based detection using Harris corner algorithm for retrieving text and caption from image. By Edge detection method we remove unwanted information of the image without affecting its original structural properties.From the corner gives the area in which more gradient is present in more than one direction will be found out. This helps to detect the edges and corners of the texts and captions on the image. We finally convert the detected characters to speech using Text-To-speech (TTS) synthesizer.

KEYWORDS: **Harris Corner Algorithm, Text-To-Speech synthesizer.**

R.MADHAN
Under Graduate Student, Department of Information Technology, Saveetha School of Engineering, Saveetha University, Chennai,

***Corresponding author**

I. INTRODUCTION

The main objective of this work is to convert the text from an image and to speech which will be very helpful for the people with low eye perception. with the help of this technology they can easily overcome their inabilities.

II. Existing SystemIn the existing system, we can use for detecting the text present in an image is considered to be complicated process due to complications like appearance of text colour variation. The conventional methods are grouped by categories like, Surface Based methods, fixed constituent based methods, and edge based method.

Problem Statement

- It is not user friendly.
- Time consuming process is high.

Proposed ystem

The proposed text detection and conversion to speech approach, we specially have industrialized a unique approach to identify text and footers in a still image. The procedure is,

1. The major feature of a image is considered to be the corners which is fundamentally robust and more salient.

2. Subsequently extraction of the curve points we go to the feature description, In order to make the system take a desired decision to accept the text from the regions we must enumerate the properties like shape of the specific regions containing.

3. Text to speech conversion, once the process of text extraction is completed we convert the extracted information say text, characters to next and final stage that is speech conversion using the text to speech synthesizer (TTS) which actually speaks the contents present in the image region. This will be useful for identifying the words, sentence for people with low eye perception.

Advantages

- Here we can implement new techniques is called neural network, and apply the Harris corner algorithm.
- Time consuming process is slow to compare the previous work
- It is user friendly.

III. MODULES DESCRIPTION

- Edge Detection
- Feature Extraction
- Text to speech

EDGE`DETECTION

The edges present in the region has to be detected so as the system can conclude and admit the text regions. We were initially converted the query image to avoid the content and color variations on the text regions over the still image if there are contrast line, spaces can be extracted by a ridge detector. Where there might be diminutive pixels in varying color combination else ageless contextual. Therefore for a line, it's typically each lines has one edge on either side of the line.

IV. FEATURE EXTRACTION

The text regions can be described as follows

- Region,
- Saturation,
- Orientation,
- Aspect ratio and
- Position.

By converting into a binary image the corner detected we can use the new techniques for neural network. To apply the harris corner algorithm to implementing the feature description by selecting the text area.

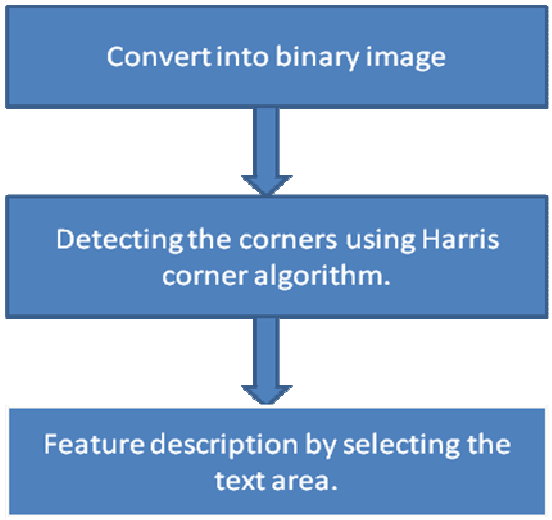

Figure 1
Feature extraction

Text To Speech Conversion

The extracted text from the query image is converted into speech with natural voice output of the extracted text. In conventional method the text to speech conversion suffers the originality and audio quality here in the proposed work uses the text to speech synthesizer which produces the uniqueness similar to the human voice.

Flow Chart of Project

The following figure shows the complete flow of the project and it is very much needed to be handled carefully to extract all of the text that is present in the image and produce the corresponding text onto an audio output. The image is being taken and given as an input for processing. The image is being processed individually so that the complete process is successful when all of the pixels are being processed. The edges are being detected so that all of the edges of the frames are found and the text can be separated from that of the frame very easily. The edge detection process is the most interesting and most important thing needed to be done with proper care so that no confusion can take place at any point of time. After detecting the edge only the sharpness of the image can be found out and then the further process can be carried out successfully.

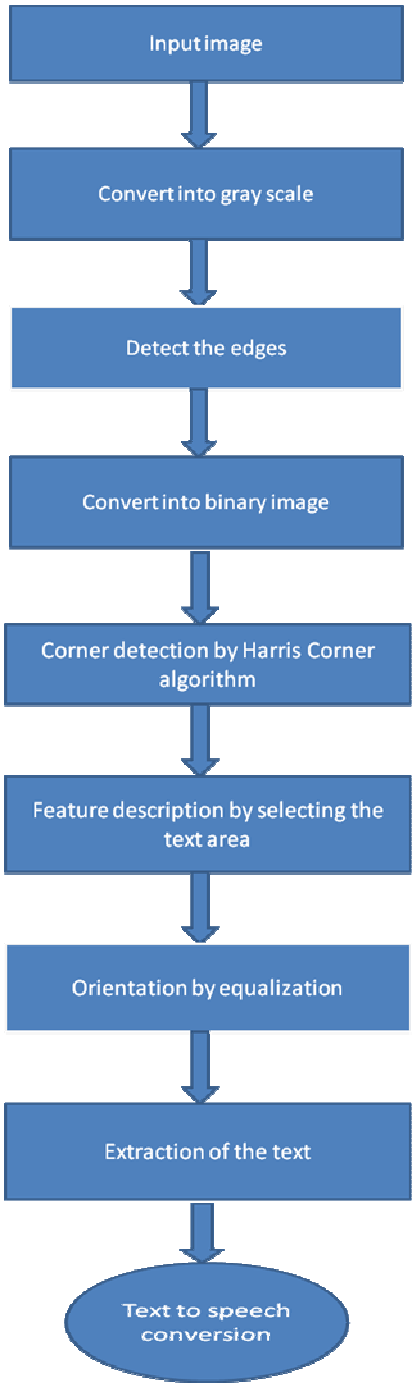

Figure 2
Flow Chart for Extraction and voice conversion

The features of the text from that of the others such as images, logos, etc can be differentiated by the properties of the text such as

- Area
- Aspect Ratio
- Saturation
- Position
- Orientation

With these properties the text can be differed. Orientation of the text can be found from that of the single frame so that the extracted text can be stored in the database. The extracted texts are shown as characters in a new pop up window which precisely shows only the characters present in the image which we took. Finally the characters are converted to speech using the speech synthesizer.

V. SYSTEM ARCHITECURE

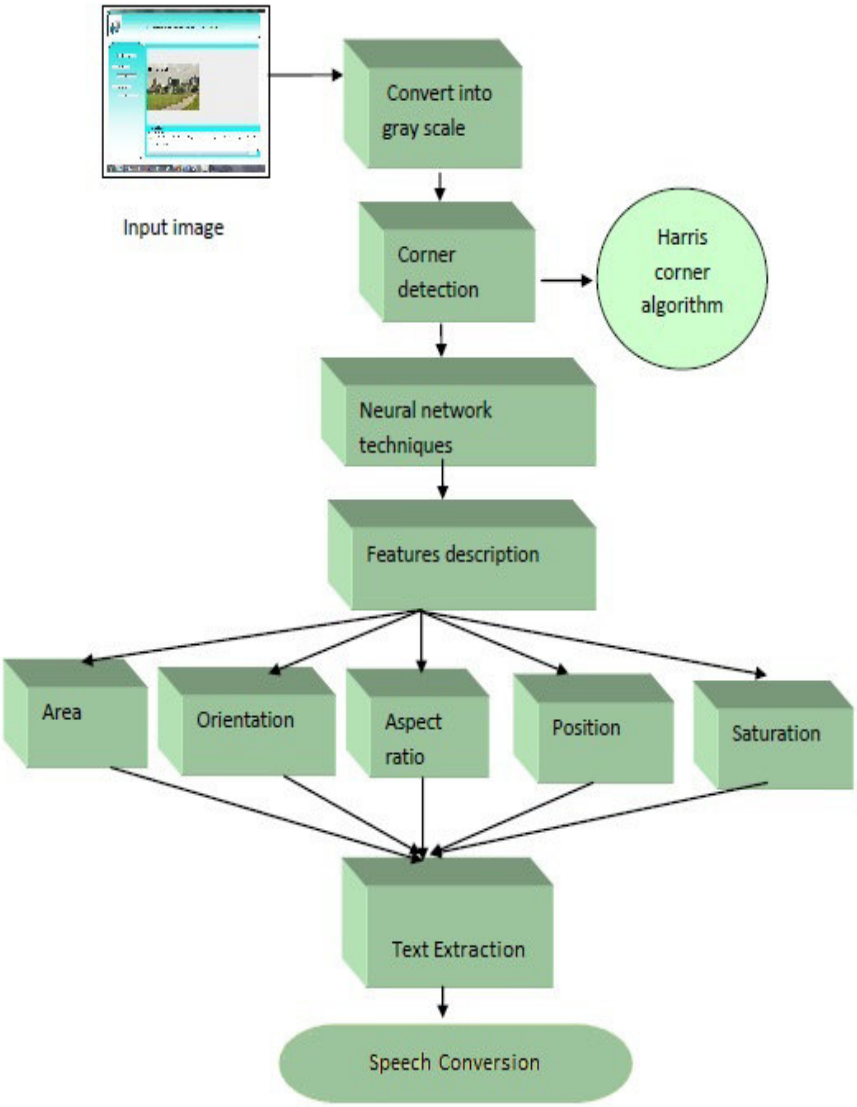

Figure 3
System Architecture

Algorithm 1: To extract text from an image

Input: Image file is given as an input where the region contains text.

Output: The image undergoes the following process to extract twv she text and convert the text into voice.

Process

Step1: The query image is persuaded into gray scale image.
Step2: The Harris corner algorithm is applied over the grayscale image to attain the corners from the regions.
Step3: The corner from the image is smoothened to extract the perfect corners and edges.
Step4: The extracted text is converting into speech using TTS algorithm.

```
for(int iy=0;iy<w;iy++)
{
for(int jy=0;jy<h;jy++)
{
clrww[iy][jy]= ex.image.getRGB(iy,jy);
t1.setRGB(iy,jy,clrww[iy][jy]);
}
}
```

VI . EXPERIMENTAL ANALYSIS

In this window the query image is taken and processed to get the text from the image.

Figure 3
Query image

In this process the original image is converted into the gray scale image and processed to avoid the text and color variation.

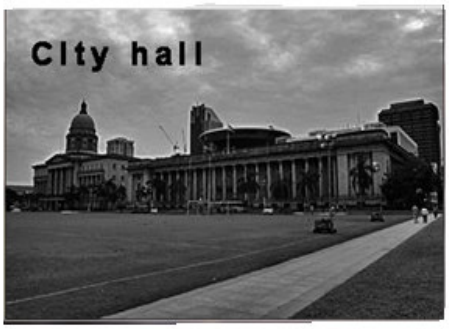

Figure 3.1
Gray scale of query image

After this process the image is divided into pixel by pixel to accquire the originality of text.

Figure 3.2
Edge detection

The image is smoothened further for producing the accurate content of the text.

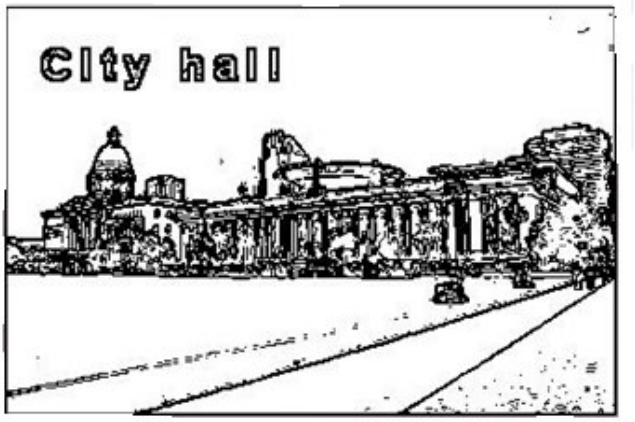

Figure 3.3
Smoothened image

Here the content is sliced from the original text region.

Figure 3.4
Sliced text region

Finally the characters are displayed and then it is converted into speech.

Figure 3.5
Output of the above process

VII. CONCLUSION

We have industrialized an impulsive text and caption recognition system from a still image which converts the extracted text to speech considering the fundamental feature as corner points of a text, characters and captions. This system detects character, text and captions with high accuracy and efficiency by giving a clear speech output of the extracted text. In conjunction with this concept of text detection based on the corner points this ideas can be adaptable to different application and the future development of the system can be implemented on video as well by taking frames from the video.

REFERENCES

1. Xu Zhao, Kai-Hsiang Lin, Yun Fu, Member, IEEE, Yuxiao Hu, Member, IEEE, Yuncai Liu, Member, IEEE, and Thomas S. Huang, Life Fellow, IEEE," MARCH 2011Text From Corners: A Novel Approach to Detect Text and Caption in Videos, IEEE Transactions on Image Processing, VOL. 20, NO. 3.
2. Wonjun Kim and Changick Kim, Member, IEEE, "A New Approach for Overlay Text Detection and Extraction From Complex Video Scene", IEEE Transactions on Image Processing, VOL. 18, NO. 2, FEBRUARY 2009.
3. K. Kim, K. Jung, and J. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 12, pp. 1631–1639, Dec. 2003.
4. X. Tang, X. GAO, J. Liu, and H. Zhang, "A spatial-temporal approach for video caption detection and recognition," IEEE Trans. Neural Netw., vol. 13, no. 4, pp. 961–971, Jul. 2002.
5. 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 Trans. Pattern Anal. Mach. Intel., vol. 22, no. 12, pp. 1349–1380, Dec. 2000.
6. Y. Zhong, H. Zhang, and A. K. Jain, "Automatic caption localization in compressed video," IEEE Trans. Pattern Anal. Mach. Intel., vol. 22, no. 4, pp. 385–392, Apr. 2000.
7. H. Li, D. Doermann, and O. KIA, "Automatic text detection and tracking in

digital video," IEEE Trans. Image Proces., vol. 9, no. 1, pp. 147–156, Jan. 2000.

8. V. Wu, R. Manmatha, and E. M. Riseman, "Textfinder: An automatic system to detect and recognize text in images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 11, pp. 1224–1229, Nov. 1999.
9. Y. A. Aslandogan and C. T. Yu, "Techniques and systems for image and video retrieval," IEEE Trans. Knowl. Data Eng., vol. 11, no. 1, pp. 56–63, Jan./Feb. 1999.
10. Jun Ohya, Akio Shio, and Shigeru Akamatsu," Recognizing Characters in Scene Images," IEEE Transactions on Pattern Analysis and Machine intelligence. VOL. 16, NO. 2, FEBRUARY 1994.
11. K.Jung,I.Kim and A.K.Jain,"Text information extraction in images and video:A Survey,"Pattern Recognition,vol.37,no.5,pp 977-997,2004.
12. Swati Talesara, Hemant A. Patil, Tanvina Patel, Hardik Sailor and Nirmesh Shah "A Novel Gaussian Filter-based Automatic Labeling of Speech Data for TTS System in Gujarati Language" 2013 International Conference on Asian Language Processing
13. Safia Shaik, Dr.Y.Padmasai, V.Naveen Kumar "DEVELOPMENT OF TELUGU TEXT TO SPEECH SYSTEM USING OMAP 3530" 978-1-4799-2845-3/13/$31.00 ©2013 IEEE.
14.]Ramani B, Actlin Jeeva M P, Vijayalakshmi P, Nagarajan T "Voice Conversion-Based Multilingual to Polyglot Speech Synthesizer for Indian Languages" 978-1-4799-2827-9/13/$31.00 ©2013 IEEE.
15. Yan-You Chen1, Yu-Wei Bai1, Chun-Yu Tsai2, Jhing-Fa Wang1,3, Bo-Wei Chen "Voice-Customizable Text-To-Speech for Intelligent Home-Care System" 978-1-4673-5936-8/13/$31.00 ©2013 IEEE