Interactive Scene Analysis

ADHISH BHATKAR, NOIDA FERNANDES, SURAJ GURAV, VINAYA WATE
UNDER PROF. SUPRAVA PATNAIK

Abstract: It is very difficult for the blind and visually impaired to cope up with the world in their daily environment. Even though technology has increased tremendously there have not been specific and helpful gadgets or applications to assist them .In this paper we have put forward the idea of presenting an application which can extract and recognize the character from the image, which is clicked by the camera of an assistive device, which is a hand held object and produce an audio output. Thus it is an easily available and efficient system that will be able to extract and recognize each character independently and efficiently.

Index Terms-- Unaided, Assistive devices, Handheld objects, Text extraction, Character recognition

1. INTRODUCTION

1,100,000 people are legally blind in the US. Approximately 1 in 247 or 0.40% or 1.1 million people in USA. India has the largest blind population in the world [1]. Text detection and recognition in natural images are popular yet unsolved problems in modern technology.

The term "scene analysis" refers to the semantic analysis of image data, often recorded outdoors. A simple example is traffic sign detection[13]. Often, however, they are recordings from different sources. One of the largest projects of the image processing department is the automatic detection of objects photographed with different mobile phones[2]. For the purpose of the largely parameter-free recognition of different objects, current research algorithms are adapted as well as new algorithms developed. For many years machine learning has been an integral part of many projects and research activities in the department. In the field of surface inspection, hybrids from the "classical" parameterizable methods (filters, morphology, edge detectors) and learning approaches are increasingly being used. Learning methods, such as "deep learning", require a large number of annotated data, which in an industrial project is usually neither affordable nor practical[4]. For this reason, it is necessary to model assumptions about the objects, you want to locate, and to use this modeling as a partial input for automated procedures. Because of this special data situation, the department's learning approaches are almost always model-based (model-based machine learning).

2. LITERATURE REVIEW

Scene text recognition has generated significant interest from many branches of research. While it is now possible to achieve extremely high performance on tasks such as digit recognition in controlled settings , the task of detecting and labeling characters in complex scenes remains an active research topic. However, many of the methods used for scene text detection and character recognition are predicated on cleverly engineered systems specific to the new task. For text detection, for instance, solutions have ranged from simple off-the-shelf classifiers trained on hand coded features to multi-stage pipelines combining many different algorithms[7]. Common features include edge features, texture descriptors, and shape contexts. Meanwhile, various flavors of probabilistic model have also been applied, folding many forms of prior knowledge into the detection and recognition system. On the other hand, some systems with highly flexible

3. EXISTING TECHNOLOGY

3.1. Portable Camera-Based Assistive Text and Product Label Reading from Hand-Held Objects for Blind Persons

fig. 1. Portable device

In this project a camera-based assistive text reading framework to help blind persons read text labels and product packaging from hand-held objects in their daily lives. To isolate the object from cluttered backgrounds or other surrounding objects in the camera view, first propose is an efficient and effective motion- based method to define a region of interest (ROI)[6]. This method extracts moving object region by a mixture-of-Gaussians-based background sub- traction method. In the extracted ROI, text localization and recognition are conducted to acquire text information.

To automatically localize the text regions from the object ROI, the propose of text localization algorithm by learning gradient features of stroke orientations and distributions of edge pixels in an Adaboost model. The figure below shows the proposed work of this project. Which consists of the following:

- Scene capturing device (a camera attached with a pair of spectacles).
- Data processing device (a personal computer).
- Audio output device (earphones).

The drawbacks of this project are as follows :

- Not portable
- Inconvenient
- Longer processing time

3.2. Text Extraction and Recognition from Image using Neural Network.

Extraction and recognition of text from image is an important step in building efficient indexing and retrieval systems for multimedia databases[5]. The primary objective of this project is to make an unconstrained image indexing and retrieval system using neural network. HSV based approaches has been adopted for color reduction. This approach show impressive results. The extraction of a set of features from each ROI for that specific color plane and use them further in a feature-based classifier to determine if the ROI contains text or non-text blocks. The blocks identified as text are next given as input to an OCR.

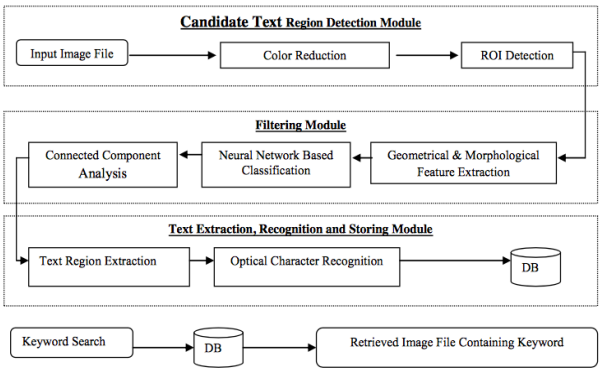


fig. 2. Text region detection module

The OCR output in the form of ASCII characters forming words is stored in a database as keywords with reference for future retrieval .The block diagram of text detection, extraction, recognition and storing is shown below along with the Block diagram of video retrieval based.

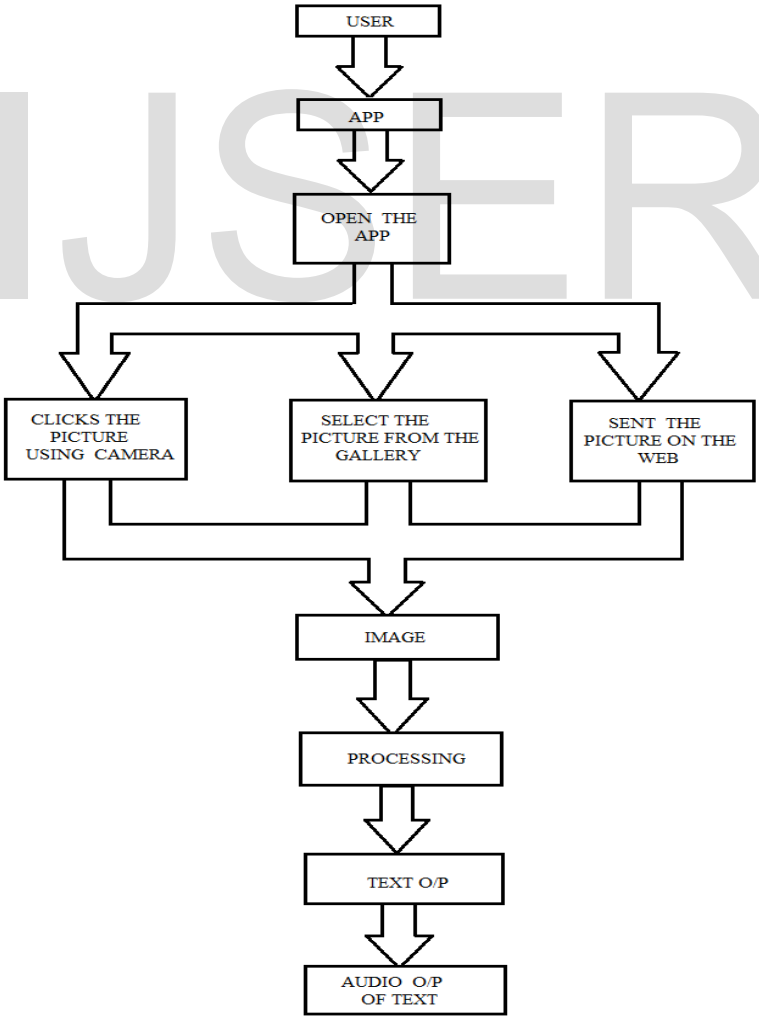
The major drawbacks of this projects are as follows:
- Slower speed due to compressed domain processing.
- System cannot detect non-horizontally text in an image.
- Improvement for better tracking of text with complex motion.
- Recognition accuracy is poor for text with complex background.
- A more accurate OCR is needed to improve the quality of retrieval further.

4. PROPOSED METHDOLOGY

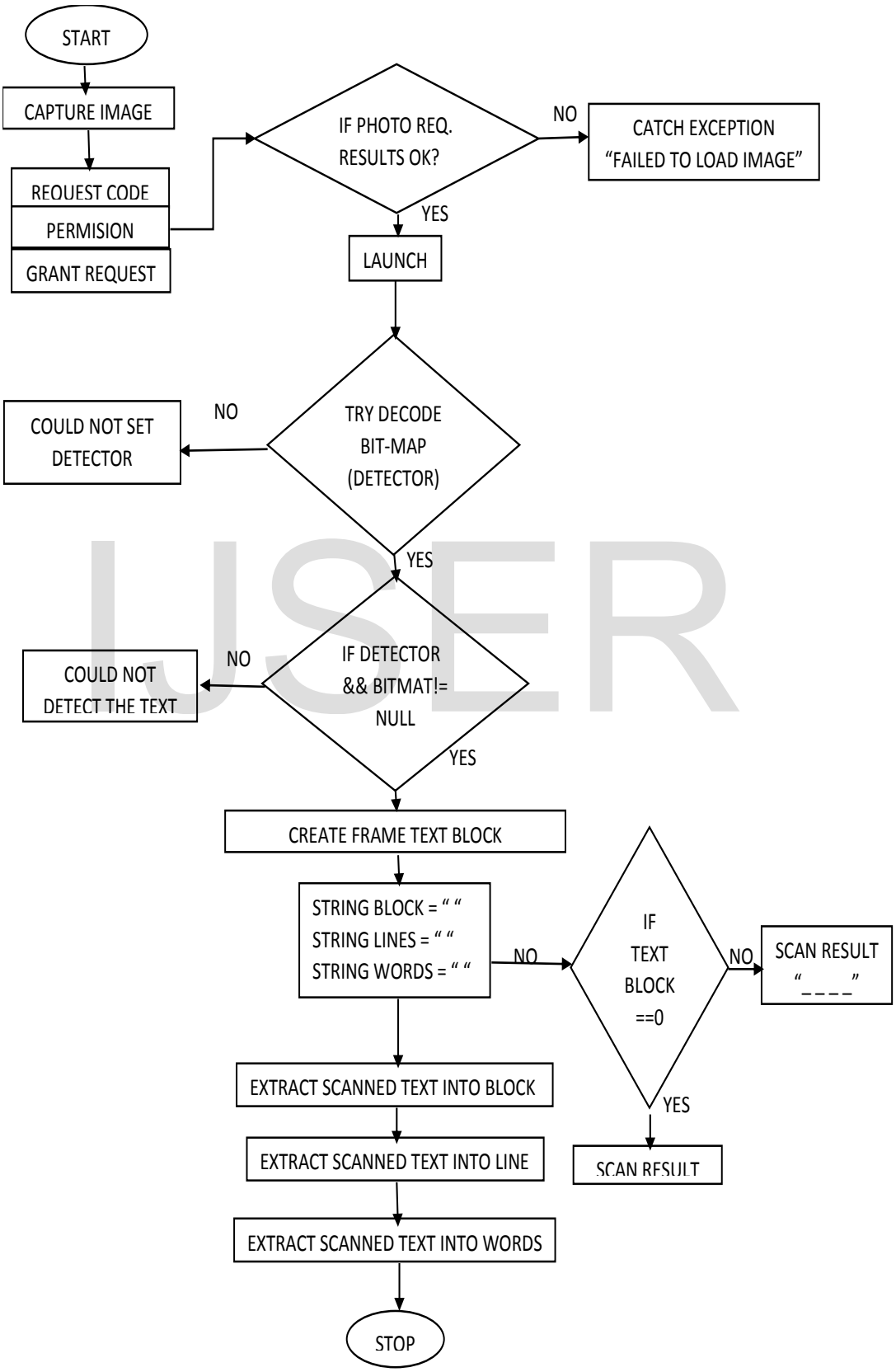
4.1 Process

As the user opens the application the screen pops up with various options such as:
- CAMERA: Here we can click the image using the camera option.
- GALLERY: This option enables the user to select the image from the system's gallery.
- WEB BROWSER: This option allows the user to surf the web in order to select the image of his/her interest.

As the image is selected the it is processed in the application due to which the user gets the text extracted from that respective image. Later on the audio output is also obtained of the extracted text.



4.2 Flow Chart



5. RESULTS

The results obtained are available in three various formats:

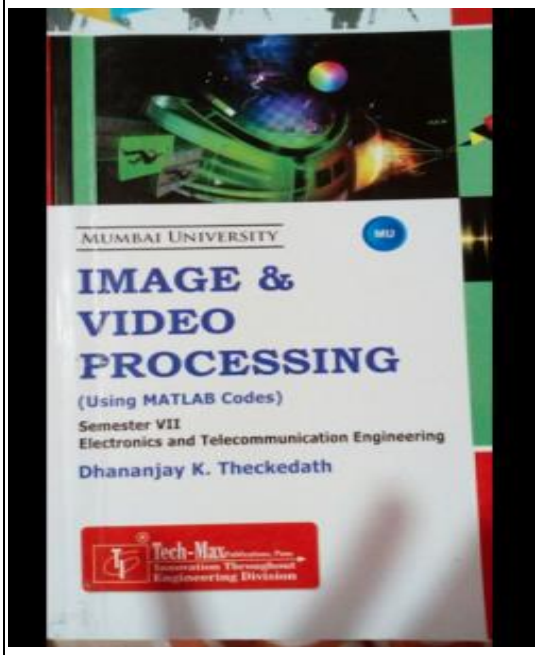
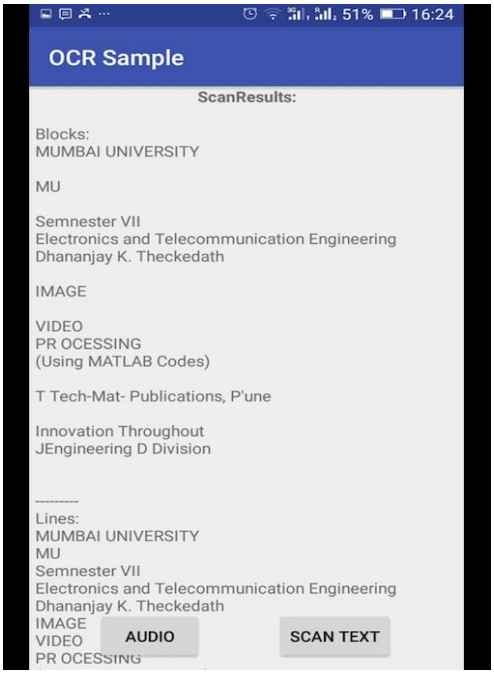
Blocks- This format is useful to locate the position of the text within the image.

Lines- It is widely applicable in the language processing domain.

Words- This is suitable for extracting information like expiry date, brand name. It can also be used for Emotion analysis.

In the next phase text along with the objects, along with the facial expression will be detected and recognized which is the future scope of this paper.

Sr No.	Scan image	Outputs
1.		

2.	 MUMBAI UNIVERSITY IMAGE & VIDEO PROCESSING (Using MATLAB Codes) Semester VII Electronics and Telecommunication Engineering Dhananjay K. Theckedath Tech-Max	 16:24 OCR Sample ScanResults: Blocks: MUMBAI UNIVERSITY MU Semnester VII Electronics and Telecommunication Engineering Dhananjay K. Theckedath IMAGE VIDEO PR OCESSING (Using MATLAB Codes) T Tech-Mat- Publications, P'une Innovation Throughout JEngineering D Division Lines: MUMBAI UNIVERSITY MU Semnester VII Electronics and Telecommunication Engineering Dhananjay K. Theckedath IMAGE VIDEO PR OCESSING AUDIO SCAN TEXT
3.	 Amazone Garamond Staccato Avant Garde Helvetica Times Chancery Kidprint University Edwardian Nuptual Vladimir Flemish Optima Viraldi Goudy Slogan Wedding	 16:18 OCR Sample ScanResults: Blocks: THa emtar Goudy Cobunrtan Pirdh ilebing Garamond Chancery. Kidprint Lriwersty Optima Avant Garde Helvetica Times Kanbual dia Stcuan Lines: THa emtar Goudy Pirdh ilebing AUDIO SCAN TEXT

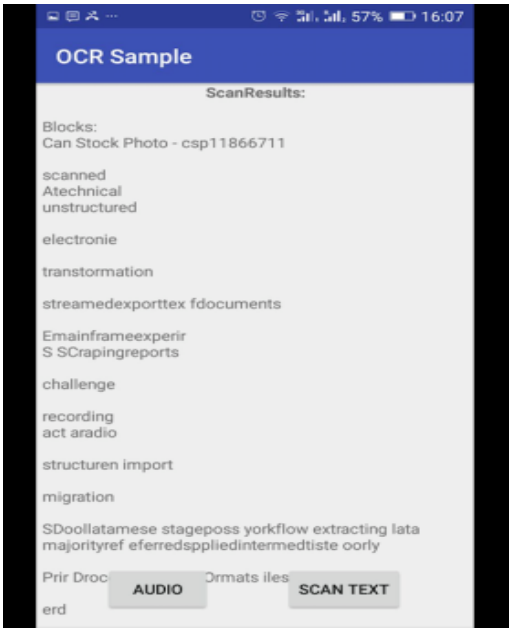
4.		

Table. 1.Result

Here we can see that for different orientations including various fonts and font-size max result is obtained.

6. CONCLUSION AND FUTURE SCOPE

In this paper we have discussed about the problems faced by the unaided civilians and how they had to cope up with their day-to-day life. So in order to overcome their difficulties we have proposed an application "***I.NTERACTIVE SCENE ANALYSIS***" where in the first phase is **text detection**. Here the text is extracted from the image, detected and then converted into audio using talk back feature in order to assist the blind.

In the next phase text along with the objects, along with the facial expression[11],[12] will be detected and recognized which is the future scope of this paper.

6.1 Challenges faced during Text Detection

Brightness:The amount of light falling on the object determines the brightness of the image. Light falling on any object is never constant during day time. And if the text to be detected is placed under a shadow of another massive object then text detection
becomes a challenging task to be performed.

This is illustrated in the fig. below.

fig. 3.Shadowed Image

Various Orientation: This is mainly concerned w.r.t. the alphabets. The image can have alphabets in various manner or orientation. The system must be capable of handling such difficulties. As shown in the fig. the character "V" may be in various forms

fig. 4. Various Orientation

Diverse Font and Font Sizes: There are various fonts and font sizes as shown in the figure aside which is a difficult task for a system.

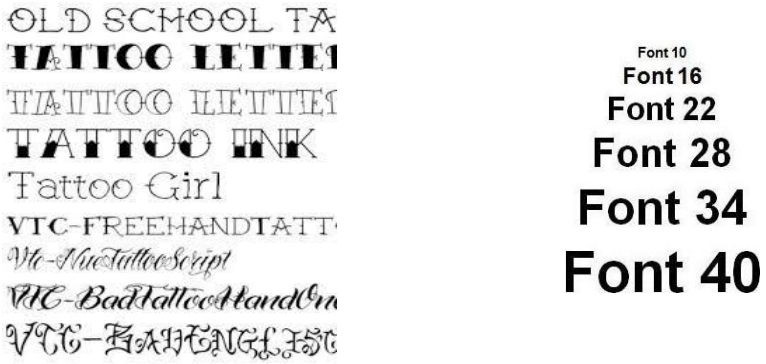


fig. 5. Different fonts and font size

REFERENCES

[1] World Health Organization. (2009). 10 facts about blindness and visual impairment [Online].Available www.who.int/features/factfiles/blindness/blindness_facts/en/index.html

[2] Advance Data Reports from the National Health Interview Survey (2008). [Online]. Available: http://www.cdc.gov/nchs/nhis/nhis_ad.html

[3]International Workshop on Camera-Based Document Analysis and Recognition (CBDAR 2005, 2007, 2009, 2011).[Online]. Availablehttp://www.m.cs.osakafuu.ac.jp/cbdar2011/

[4] X. Chen and A. L. Yuille, "Detecting and reading text in natural scenes," in *Proc. Comput. Vision Pattern Recognit.*, 2004, vol. 2, pp. II-366–II-373.

[5] X. Chen, J. Yang, J. Zhang, and A. Waibel, "Automatic detection and recognition of signs from natural scenes," *IEEE Trans. Image Process.*, vol. 13, no. 1, pp. 87–99, Jan. 2004.

[6] D. Dakopoulos and N. G. Bourbakis, "Wearable obstacle avoidance electronic travel aids for blind: A survey," *IEEE Trans. Syst., Man, Cybern.*, vol. 40, no. 1, pp. 25–35, Jan. 2010.

[7] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," in *Proc. Comput. Vision Pattern Recognit.*, 2010, pp. 2963–2970.

[8] Y. Freund and R. Schapire, "Experiments with a new boosting algorithm," in *Proc. Int. Conf. Machine Learning*, 1996, pp. 148–156.

[9] Bourne RRA, Flaxman SR, Braithwaite T, Cicinelli MV, Das A, Jonas JB, et al.; Vision Loss Expert Group. Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. Lancet Glob Health. 2017 Sep;5(9):e888–97.

[10] Portable Camera-Based Assistive Text and Product Chucai Yi, Student Member, IEEE, Yingli Tian, Senior Member, IEEE, and Aries Arditi IEEE/ASME TRANSACTIONS ON MECHATRONICS, VOL. 19, NO. 3, JUNE 2014

[11] Example-Based Object Detection in Images by Components Anuj Mohan, Constantine Papageorgiou, and Tomaso Poggio, Member, IEEE

[12]Techniques for Object Recognition in Images and Multi-Object Detection Khushboo Khurana, Reetu Awasthi

[13]Wikipedia, the free encyclopedia