



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 1, January 2015

## Text Extraction from Natural Scene Images and Conversion to Audio in Smart Phone Applications

M. Prabakaran<sup>1</sup>, K. Radha<sup>2</sup>

M.E Student, Department of Computer Science and Engineering, Muthayammal Engineering College, India<sup>1</sup>

Assistant Professor, Department of Computer Science and Engineering, Muthayammal Engineering College, India<sup>2</sup>

**ABSTRACT:** Extracting text character from natural scene images is a challenging problem due to differences in text style, font, size, orientation, alignment and complex background. The text data present in images and video contain certain useful information for content-based information indexing and retrieval, sign translation and intelligent driving assistance. In scene text extraction, adjacent character grouping and character stroke orientation method is performed to search for image regions of text strings. Our proposed system is extracting a text and the extracted text information into audio. The Smart mobile phone application is used to show the effectiveness of our proposed method effectively.

**KEYWORDS:** Scene Text Extraction, Character Stroke Orientation, Smart Phone Application

### I. INTRODUCTION

Extracting text from images or videos is an important problem in many applications like document processing, image indexing, video content summary, video retrieval, video understanding. In natural scene images and videos, text characters and strings usually appear in nearby sign boards and hand-held objects and provide significant knowledge of surrounding environment and objects. Natural scene images usually suffer from low resolution and low quality, perspective distortion and complex background [1].

Scene text is hard to detect, extract and recognize since it can appear with any slant, tilt, in any lighting, upon any surface and may be partially occluded. Many approaches for text detection from natural scene images have been proposed recently. To extract text information by mobile devices from natural scene, automatic and efficient scene text detection and recognition algorithms are essential. The main contributions of this paper are associated with the proposed two recognition schemes. Firstly, a character descriptor is proposed to extract representative and discriminative features from character patches. It combines several feature detectors (Harris-Corner, Maximal Stable Extremal Regions (MSER), and dense sampling) and Histogram of Oriented Gradients (HOG) descriptors [5]. Secondly, to generate a binary classifier for each character class in text retrieval; we propose a novel stroke configuration from character boundary and skeleton to model character structure.

The proposed method combines scene text detection and scene text recognition algorithms. By the character recognizer, text understanding is able to provide surrounding text information for mobile applications, and by the character classifier of each character class, text retrieval is able to help search for expect objects from environment. Similar to other methods, our proposed feature representation is based on the state-of-the-art low-level feature descriptors and coding/pooling schemes. Different from other methods, our method combines the low-level feature descriptors with stroke configuration to model text character structure. Also, we present the respective concepts of text understanding and text retrieval and evaluate our proposed character feature representation based on the two schemes in our experiments. Besides, previous work rarely presents the mobile implementation of scene text extraction, and we transplant our method into an Android-based platform.

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 1, January 2015

## II. RELATED WORK

Current optical character recognition (OCR) systems can achieve almost perfect recognition rate on printed text in scanned documents, but cannot accurately recognize text information directly from camera-captured scene images and videos. Lu et al. [3] modeled the inner character structure by defining a dictionary of basic shape codes to perform character and word retrieval without OCR on scanned documents. Coates et al. [5] extracted local features of character patches from an unsupervised learning method associates with a variant of K-means clustering, and pooled them by cascading sub-patch features. In [8], a complete performance evaluation of scene text character recognition was carried out to design a discriminative feature representation of scene text character structure. Weinman et al. [7] combined the Gabor-based appearance model, a language model related to simultaneity frequency and letter case, similarity model, and lexicon model to perform scene character recognition. Neumann et al. [1] proposed a real time scene text localization and recognition method based on extremal regions Smith et al. [2] built a similarity model of scene text characters based on SIFT, and maximized posterior probability of similarity constraints by integer programming. Mishra et al. [9] adopted conditional random field to combine bottom-up character recognition and top-down word-level recognition.

## III. PROPOSED DESIGN

Text detection and recognition used to detect text in complex background images. It takes text image as input and then applying preprocessing methods on it to remove noise from image by converting color image to gray ,binarization which helps to efficient and accurate text identification from image which is input to OCR, within preprocessing if some part text data will loss them by thinning and scaling is performed by connectivity algorithm. Then we get connected text character from image. Then text recognition is done. The proposed framework is divided into three stages. Here applied text detection and text recognition to the image and recognize. The text detection uses to quickly extract text region in images with a very less false positive rate. To provide the recognition for accurate result we proposed system to test the text image is segmented, assuming a different number of classes in the image each time.

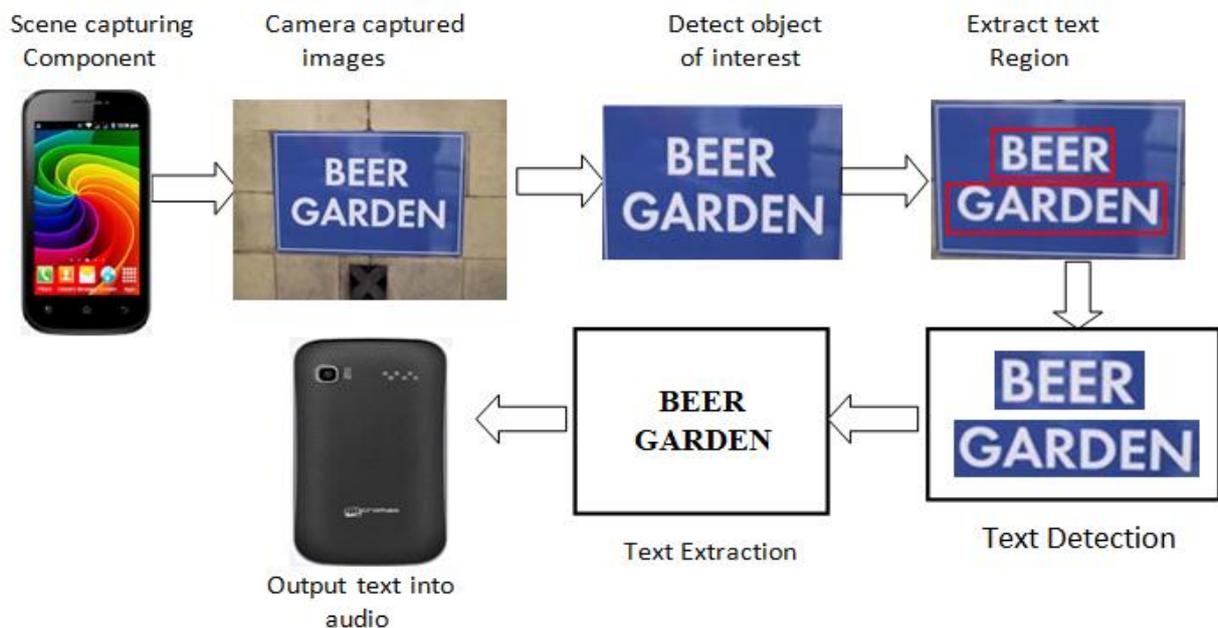


Fig1. The flowchart of the proposed method.

It is a demanding problem to automatically localize objects and text Region of Interest from captured images with cluttered backgrounds, because text in natural scene images is most likely surrounded by various surroundings outlier noise and text characters usually appear in different fonts and colors. For the content orientations, this thesis assumes

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 1, January 2015

that text strings in scene images keep in the section of vertical locations. Several algorithms have been developed for localization of text area in scene images. We can divide them into categories component Based and operations Based.

To recognize and extract text from difficult backgrounds with multiple and variable text patterns, here propose a text localization algorithm that combines area-based layout analysis and horizontal-based text classifier preparation, which define characteristic maps based on stroke orientations and boundary distributions. To generate delegate and discriminative text features to decide text characters from environment outliers.

## A. TEXT DETECTION:

The text detection stage search for to detect the occurrence of content in a camera captured natural scene images. Because of different font, highlights, different cluttered background image alteration and demeaning correct and quick text detection in scene images is still difficult task. The approach uses a character descriptor to fragment text from an image. Initially content is detected in multi size images using edge based system, morphological Function and projection report of the image. These detected text area are then confirmed using descriptor and wavelet features. The algorithm is strong when difference in style, size of font and color. Vertical edges with a predefined pattern are used to detect the edges, then grouping vertical boundaries into text area using a filtering process.

### • *Adjacent text grouping process*

The text information usually appears in text strings composed of several character members in similar sizes rather than single character, and text strings are normally in approximately horizontal alignment. The adjacent character grouping method calculates the sibling groups of each character candidate as string segments and then merges the intersecting sibling groups into text string. To model the boundary size and location of a text string, a bounding box is assigned to each boundary in a color layer.

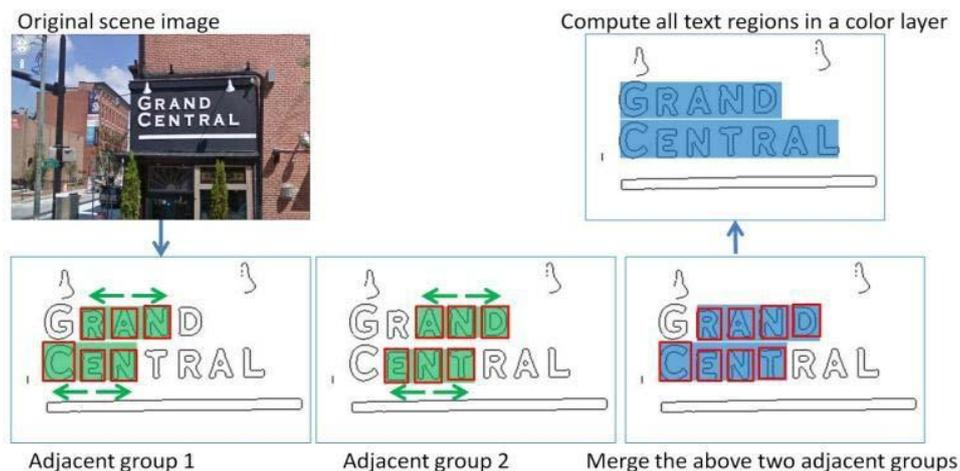


Fig. 2. The adjacent text grouping method in natural scene image.

The red box denotes bounding box of a boundary in a color layer. The green regions in the bottom left two figures represent two adjacent groups of consecutive neighboring bounding boxes in similar size and horizontal alignment. The blue regions in the bottom-right figure represent the text string fragments, obtained by merging the overlapping adjacent groups. For each bounding box, we search for its siblings in similar size and vertical locations. If several neighbor bounding boxes are obtained on its left and right and to join all these involved boxes into a areas. This section contains a fragment of text sequence. Finally this technique is to calculate all text sequence fragments in this color layer, and merge the string fragments with intersections.

In order to extract text strings in slightly non-horizontal orientations. To search for feasible typescript of a text within a practical range of horizontal direction. When calculate approximately horizontal alignment, we do not require all the characters exactly arrange in a line in horizontal orientation, but allow some differences between nearby characters that are give into the similar string. In our system we set this series as  $\pm\pi/8$  degrees comparative to the horizontal line. This

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 1, January 2015

range could be set to be better but it would bring in more false positive strings from environment. Proposed image into text detection algorithm can hold demanding lettering variations, as long as the text has sufficient resolutions, such as newspaper heading.

## B. Text Extraction

To detect and extracting the text from camera captured natural scene images. Carefully extract the text from nearby object held by the blind person from the cluttered conditions. Text finding is used to obtain text containing image area then text identification to transform image-based information into readable text. This step refers to classify the characters as they are in original image. This is done by multiplying resultant figure with binary converted novel image. Final result is the white text in black background dependent on the novel image.

### • *Text Character Stroke Configuration*

Text characters consist of strokes with constant or variable orientation as the basic structure. Here, we propose a new type of feature, stroke orientation, to describe the local structure of text characters. From the pixel based analysis, stroke direction is vertical to the slope orientations at pixels of stroke borders. To model the text structure by stroke orientations, we propose a new operator to map a gradient feature of strokes to each pixel. It extends the local structure of a stroke edge into its neighborhood by gradient of orientations. We use it to develop a attribute map to analyze total arrangement of text characters.

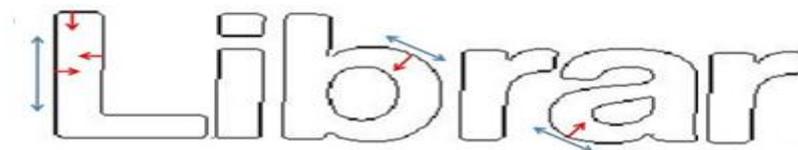


Fig.3. A text stroke configuration at pixels of stroke borders.

Blue arrows denote the stroke configuration at the sections and red arrows denote the gradient configuration at pixels of stroke borders. In order to locate stroke exactly, stroke is redefined in our algorithm as outline points within character part with constant width and point of reference. A character can be corresponding to as a set of linked strokes with specific configuration which includes the number, letters, alignment and length of the strokes. Here, the structure map of strokes is defined as stroke configuration. In a character class, although the character instances appear in different sizes, styles, and fonts, the stroke orientation is always constant. For example, character 'B' is always a vertical stroke with two arc strokes in any model. Therefore for each of the 62 character classes, we can estimate a stroke configuration from training patches to describe its basic structure. The Synthetic Font Training Dataset proposed by [7] is employed to obtain stroke configuration. This dataset contains about 67400 character patches of synthetic English letters and digits in various fonts and styles, and we select 20000 patches to generate character patches.

It covers all the 62 classes of characters. Each character image is normalized into  $128 \times 128$  pixels with no antialiasing. In estimating stroke configuration, character boundaries and skeletons are generated to extract stroke-related features, which are used to compose stroke configuration. The implementation contains three main steps as follows. Firstly, given a synthesized character patch from the training set, we obtain character boundary and character skeleton by applying discrete contour evolution (DCE) and skeleton pruning on the basis of DCE [9].

## C. Extracted Text into Audio

The mobile speaker is to inform the user of extracted text codes in the type of speech or audio. Mobile phone speaker is employed for speech output.

Text extraction is performed by the current OCR earlier to crop of useful terms from the extracted text areas. A text area labels the minimum rectangular part for the place of lettering within it, so the margin of the text area links the edge border line of the text quality. On the other hand, the current system provides better performance if text section are first assigned proper margin areas and binaries to segment text characters from environment. Thus each restricted text part is enlarged by enhancing the height and width by 10 pixels respectively. We test both open and closed basis solutions exist that have APIs that allocate the ending stage of translation to letter codes.



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 1, January 2015

The recognized text codes are recorded in script files. Then we build the Text to speech to load the script files and display the audio output of text information. Users can adjust speech volume, tone and rate according to their preferences.

## IV. CONCLUSION AND FUTUREWORK

We have presented a method of scene text extraction from detected text regions, which is compatible with android mobile applications and extracted text is converted into audio. This system reads the text information in the objects and informs blind users of the extracted text information. It detects text area from natural scene image and extracts text information from the detected text regions. In image text detection, analysis of color decomposition and horizontal alignment is performed to search for image regions of text strings. This method can effectively differentiate the object of interest from background or other objects in the camera vision. Adjacent character grouping is performed to calculate candidates of text patches prepared for text classification. An Adaboost learning model is applied to localize text in camera-based images. Text extraction is used to perform word recognition on the localized text regions and transform into audio output for blind users. To model text character structure for text retrieval scheme, we have designed a novel feature representation, stroke configuration map, based on boundary and skeleton. The system demonstrates the effectiveness of our proposed method in blind-assistant applications, and it also proves that the assumptions of color uniformity and aligned arrangement are suitable for the captured text information from natural scene.

## REFERENCES

- 1.B. Epshtein, E. Ofek, and Y. Wexler, 'Detecting text in natural scenes with stroke width transform', in Proc. CVPR, pp. 2963–2970, 2010.
- 2.R. Beaufort and C. Mancias-Thillou, 'A weighted finite-state framework for correcting errors in natural scene OCR', in Proc. 9th Int. Conf. Document Anal. Recognit., pp. 889–893, 2007.
- 3.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, 2004.
- 4.A. Coates et al., 'Text detection and character recognition in scene images with unsupervised feature learning', in Proc. ICDAR, pp. 440–445, 2011.
5. N. Dalal and B. Triggs, 'Histograms of oriented gradients for human detection', in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 886–893, 2005.
- 6.T. de Campos, B. Babu, and M. Varma, 'Character recognition in natural images', in Proc. VISAPP, 2009.
- 7.R. Smith, 'An overview of the tesseract OCR engine', in Proc. Int. Conf. Document Anal. Recognit., pp. 629–633 2007.
- 8.S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong, and R. Young, 'ICDAR 2003 robust reading competitions', in Proc. Int. Conf. Document Anal. Recognit., pp. 682–687, 2003.
- 9.J. Zhang and R. Kasturi, 'Extraction of text objects in video documents: Recent progress', in Proc. 8th IAPR Int. Workshop DAS, , pp. 5–17, 2008.
- 10.Q. Zheng, K. Chen, Y. Zhou, G. Cong, and H. Guan, 'Text localization and recognition in complex scenes using local features', in Proc. 10<sup>th</sup> ACCV, pp. 121–132, 2010.

## BIOGRAPHY

**M.Prabaharan** is a M.E (CSE) Student in the Department of Computer Science and Engineering, Muthayammal Engineering College, Rasipuram, Tamil Nadu, India. He received Bachelor of Technology in Information Technology degree in 2012 from Narasu's Sarathy Institute of Technology, Salem, Tamil Nadu, India. He research interests are Image processing and Networking.

**K.Radha** is an Assistant Professor in the Department of Computer Science and Engineering, Muthayammal Engineering College, Rasipuram, Tamil Nadu, India. He received M.E (CSE) degree in 2009 from Sona College of Technology, Salem, Tamil Nadu, India. He received B.Tech (IT) degree in 2005 from Muthayammal Engineering College, Rasipuram, Tamil Nadu, India. He has been professional teaching experience in 6 years. He research interests are Medical Image Processing.