ISSN: 2321-7782 (Online)
Volume 3, Issue 4, April 2015
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Text Extraction in Android Mobile Application using Character Descriptor

Prajakta D. Sawle[1]
M.E 4th Sem Dept. of computer Science & Engineering
P. R. Patil COET, Amravati,
Maharashtra, INDIA

Prof. Chetan J. Shelke[2]
Department of Information Technology
P. R. Patil COET, Amravati,
Maharashtra, INDIA

Abstract: Images contain various types of useful information that should be extracted whenever required. A large number of algorithms and methods are proposed to extract text from the given image, and because of that a user will be able to access the text from any image. Extraction of this information involves text region detection, text localization, tracking, character extraction, enhancement, and recognition of the text from a given image. Variations in text may occur because of differences in size, style, orientation, alignment of text, and low image contrast, composite backgrounds make the problem during extraction of text. In text detection, our previously proposed algorithms are applied to obtain text regions from scene image. First, we design a discriminative character descriptor by combining several state-of-the-art feature detectors and descriptors. Second, we model character structure at each character class by designing stroke configuration maps. Our algorithm design is compatible with the application of scene text extraction in smart mobile devices. In this paper, we used cloud based API to increase the speed and accuracy of the recognition.

Keywords: -Text extraction, Text detection, Text localization, Text retrieval, OCR.

I. INTRODUCTION

Camera-based text information serves as effective tags or clues for many mobile applications associated with media analysis, content retrieval, scene understanding, and assistant navigation. 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. Text-based tags are much more applicable than barcode or quick response code [1], [2], because the latter techniques contain limited information and require pre-installed marks. Content-based image indexing refers to the process of attaching labels to images based on their content. Image content can be divided into two main categories: perceptual content and semantic content [3]. Perceptual content includes attributes such as color, intensity, shape, texture, and their temporal changes, whereas semantic content means objects, events, and their relations.

Text Extraction from image is concerned with extracting the relevant text data from a collection of images. Rapid development of digital technology has resulted in digitization of all categories of materials. Lot of resources are available in electronic medium. Many existing paper-based collections , historical manuscripts , records, books, journals, scanned document , book covers , video images, maps, manuscripts, pamphlets, posters, broadsides, newspapers,, micro facsimile, microfilms, university archives, slides and films, book plates, pictures, painting, graphic materials, coins and currency, stamps, magazines, clipping files, educational , TV programs , business card, magazines, advertisements, web pages , mixed text-picture-graphics regions etc are converted to images. These images present many challenging research issues in text extraction and recognition. Text extraction from images have many useful applications in document analysis , detection of vehicle license plate, analysis of article with tables, maps, charts, diagrams etc., keyword based image search, identification of parts in industrial automation , content based retrieval, name plates, object identification, street signs, text based video indexing, video content analysis, page segmentation, document retrieving, address block location etc [4]. However due to the variety of fonts, sizes, styles, orientations, alignment effects of uncontrolled illuminations, reflections, shadows, the distortion due to perspective projection as well as the complexity of image background, automatic Localizing and extracting text is a challenging problem [5].

Localizing text from scene images are used for navigation, assistive reading, geocoding, and content-based image retrieval. Natural scene images with text information are divided into two categories according to the complexity of the background.

1) Text characters and strings are in high resolution with a relatively simple background.

2) Text into more complex backgrounds with various natural objects, such as buildings, trees, lawns, roads [6].

II. Related Work

In Detecting Text in Natural Scenes with Stroke Width Transform is a novel image operator that seeks to find the value of stroke width for each image pixel, and demonstrate its use on the task of text detection in natural images [6]. S.Audithan et.al [7] formulated an efficient and computationally fast method to extract text regions from documents. They proposed Haar discrete wavelet transform to detect edges of candidate text regions. Non-text edges were removed using thresholding technique. They used morphological dilation operator to connect the isolated candidate text edge and then a line feature vector graph was generated based on the edge map. This method exploited an improved canny edge detector to detect text pixels. The stroke information was extracted the spatial distribution of edge pixels. Finally text regions were generated and filtered according to line features.

The unique approach of Shyamaet .al [8] projected a text segmentation technique to extract text from any type of camera grabbed frame image or video. Colour based segmentation methodology was used to link consecutive pixels in the same direction by exploiting the general text properties. Light Edge Enhancement (LEE) was used to find a set of consecutive candidate points and enhance the edge between them. Next, heavy edge enhancement (HEE) was applied to remove or reduce motion blur from camera image sequences. This helped to treat camera images and video frames in the same manner. Pan et.al [9] proposed a novel hybrid method where in a text region detector was designed to generate a text confidence map. A Local binarization approach was used to segment the text components using text confidence map. A Conditional Random Field (CRF) model was used to label components as text or non-text which was solved by minimum classification error (MCE) learning and graph cuts inference algorithm. A learning based method by building neighbouring components into minimum spanning tree (MST) and cutting off interline edge with an energy minimization model to group the text components into text lines. G. Sahooet.al [10] projected a set of sequential algorithms for text extraction and enhancement of image using cellular automata. The Luminance-based algorithm was used to convert the image in to grey scale image. Converted image have only luminosity attribute. The edge detection was performed using a 3×3 Sobel operator and it was then followed by the elimination of non-maxima and thresholding of weak edges. The edge-bounded averaging was performed through Moore neighbour hood to obtain smooth non-edge regions. The image was classified in to text based or non-text based region using constant threshold.

Sunil et.al [11] proposed a scheme for the extraction of textual areas from an image using Globally Matched Wavelet Filters (GMW) filters with Fisher classifiers .GMW filters was estimated using clustering-based technique. They have used these filters to segment the document images and classify them into text, background, and picture components. To improve the result Markov random field (MRF) based post processing had been applied.

III. System Implementation And Working

Natural scene images contain some text information which is often required to be automatically recognized and processed. Scene text may be any textual part of the scene images such as street signs, name plates. The research field of scene text recognition receives a growing attention due to the proliferation of digital cameras and the great variety of potential applications, as well. Such applications include robotic vision, image retrieval, intelligent navigation systems and applications to provide assistance to visual impaired persons.

As we know that Natural scene images usually suffer from low resolution and low quality, perspective distortion and complex background. To overcome such type of problems I introduced a method which is character descriptor. Following fig shows the flow chart of the proposed system.

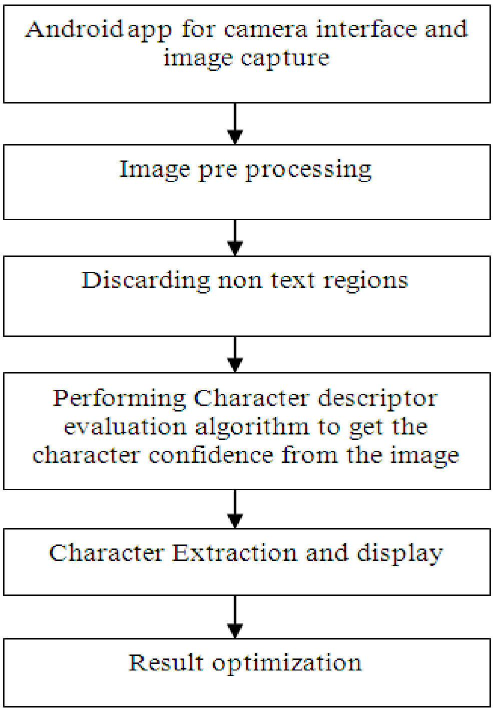

Fig.1 Flowchart of the System

a) Android app for camera interface and image capture

The first module of the system is to develop an android application for camera interfacing and image capture. The android app will capture the input scene image, on which text extraction will be performed.

b) Image preprocessing

The second module of the system is image preprocessing. The preprocessing will be done in order to get noise free and filtered image. The preprocesing will be done in two steps.

1. Converting RGB image into Grey scale image

As we know that, the image captured by camera is generally is color image i.e RGB image. The RGB color model is made up of three colors i.e red, green and blue. The name of the model comes from the initials of the three additive primary colors, red, green, and blue.The main purpose of the RGB color model is for the sensing, representation, and display of images in electronic systems, such as televisions and computers, though it has also been used in conventional photography. Before the electronic age, the RGB color model already had a solid theory behind it, based in human perception of colors. Each color have 8 pixels. The RGB image can be converted into grey scale image with the help of following formula.

Grey scale image= 0.3r + 0.59g + 0.11b

2. Applying Sobel Edge Detection algorithm

In the second step of image preprocessing , the sobel edge detection have to apply on grey scale image. The sobel edge detection algorithm is very efficient than the canny edge detection algorithm because it does not consume time as like edge detection algorithm. Compared to other edge operator, Sobel has two main advantages: Since the introduction of the average

 ISSN: 2321-7782 (Online)

factor, it has some smoothing effect to the random noise of the image. Because it is the differential of two rows or two columns, so the elements of the edge on both sides have been enhanced, so that the edge seems thick and bright.

-1	-2	-1
0	0	0
1	2	1

-1	0	-1
-2	0	2
-1	0	1

Fig.2 Sobel edge operator

These kernels are designed to respond maximally to edges running vertically and horizontally relative to the pixel grid, one kernel for each of the two perpendicular orientations. The kernels can be applied separately to the input image, to produce separate measurements of the gradient component in each orientation (call these *Gx* and *Gy*). These can then be combined together to find the absolute magnitude of the gradient at each point and the orientation of that gradient .The gradient magnitude is given by [74].

$$|G| = \sqrt{Gx^2 + Gy^2}$$

c) Discarding Non-text regions

In the third module of the system the non text regions will be discarded from the image.

d) Character Descriptor algorithm

In the fourth module the character descriptor algorithm will be apply on the image. The character descriptor algorithm is integrated with cloud based API.

1. The character descriptor is applying with the help of tess two library. The tess two library help to extract text from image. Fig.3 depicts the flowchart of proposed character descriptor. It employs four types of key-point detectors, Harris detector (HD) to extract key-points from corners and junctions, MSER detector (MD) to extract key-points from stroke components, Dense detector (DD) to uniformly extract key-points, and Random detector (RD) to extract the pre-set number of key-points in a random pattern. As shown in Fig. 3, four feature detectors are able to cover almost all representatives.

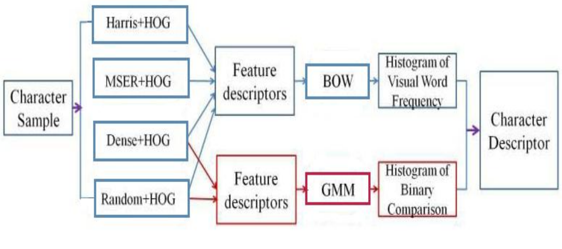

Fig 3. Flowchart of Proposed Character Descriptor

BOW: The BOW model represents a character patch from the training set as a frequency histogram of visual words. The BOW representation is computationally efficient and resistant to intra-class variations. At first, *k*-means clustering is performed on HOG features extracted from training patches to build a vocabulary of visual words. Then feature coding and pooling are performed to map all HOG features from a character patch into a histogram of visual words. We adopt soft-assignment coding and average pooling schemes in the experiments. More other coding/pooling schemes will be tested in our future work. For each of the four feature detectors HD, MD, DD, and RD, we build a vocabulary of 256 visual words. This number of visual words is experimentally chosen to balance the performance of character recognition and the computation cost. At a character patch, the four detectors are applied to extract their respective keypoints, and then their corresponding HOG features are mapped into the respective vocabularies, obtaining four frequency histograms of visual words. Each histogram has 256 dimensions. Then we cascade the four histograms into BOW-based feature representation in 256 × 4 = 1024 dimensions.

GMM: In DD and RD, keypoints are extracted from each character patch according to predefined parameters rather than character structure. In our experiments, DD generates a uniform 8 ×8 keypoint array and RD generates 64 keypoints randomly, but all character patches share the same random pattern. Therefore, the keypoints extracted by RD and DD are always located at the same positions in all character patches, as shown in Fig. 6. To describe the local feature distributions, we build a GMM over all character patches in training set. In this, each GMM contains 8 Gaussian distributions. This parameter is selected from the best results of scene character recognition.

2. The character descriptor algorithm is integrated with the help of cloud based API. The cloud based API help to increase accuracy and speed of recognition of text. ABBYY Mobile OCR Engine is a powerful software development kit which allows developers of mobile and small footprint applications to integrate highly accurate optical character recognition (OCR) technologies that convert images and photographs into manageable and searchable text. Toolkit supports the most popular mobile platforms and devices - iOS and Android.

e) Character Extraction and Display

In the fifth module, the process of character extraction and display will be done.

f) Result Optimization

In the sixth module, result optimization will be done.

Working

The user will capture the natural scene image as a input with the help of android application. After, on that image preprocessing will be done. In which the rgb image is converted into grey scale image and sobel edge detection algorithm will be applied. The image get from the second module will proceed for the next step non text regions will be discarded.After that character descriptor algorithm will be applied on the image with the help of cloud based API. The character confidence will gives the text from image and display the text as output to the user.

IV. Result Analysis

Earlier OCR technique has some drawbacks. The output of OCR technique is quite noisy and garbled. The output of the proposed system is much better than the OCR technique.

V. Conclusion And Future Scope

Thus I introduced a method or a technique which extracts the text from image. In the earlier there are many techniques which also extract text from the image, like binarisation, sliding window approach. As I proposed the system which gives better result than OCR technique. The main feature of this system is that it uses cloud based API which gives high accuracy and speed of recognition of text.

In future scope, lexicon analysis will be added to extend system to word-level recognition. This will combine scene text extraction with other techniques like content-based image retrieval to develop more useful vision based assistant system.

References

1. Y. Liu, J. Yang, and M. Liu, "Recognition of QR code with mobile phones," in Proc. CCDC, Jul. 2008, pp. 203–206.
2. E. Ohbuchi, H. Hanaizumi, and L. A. Hock, "Barcode readers using the camera device in mobile phones," in Proc. Int. Conf. Cyberworlds, Nov. 2004, pp. 260–265.
3. H. K. Kim, Efficient Automatic Text Location Method and Content-Based Indexing and Structuring of Video Database, Journal of Visual Communication and Image Representation 7 (4) (1996) 336-344.

4. C.P. Sumathi, T. Santhanam and G.Gayathri Devi,"A survey on various approaches of text extraction in images" International Journal of Computer Science & Engineering Survey (IJCSES) Vol.3, No.4, August 2012,pp. 27-42.

5. G. Rama, Mohan Babu, P. Srimaiyee, A. Srikrishna,"Text Extraction From Heterogenous images using mathematical morphology, Journal of Theoretical and Applied Information Technology © 2005 – 2010,pp. 39-47.

6. Fathima A Muhmmadali, Shibu K R,"Localising Multioriented Text in Scene Imges", Proceedings of Second IRF International Conference on 10th August 2014, Cochin, India, ISBN: 978-93-84209-43-8 pp. 102-105.

7. Audithan,,R.M.Chandrasekaran (2009), "Document Text Extraction From Document Images Using Haar Discrete Wavelet Transform",European Journal Of Scientific Research, Vol.36 No.4 , pp. 502-512.

8. Shivakumara P, A Dutta, U Pal And C L Tan(2010), "A New Method For Handwritten Scene Text Detection In Video", International Conference On Frontiers In Handwriting Recognition, pp: 16-18.

9. S. A. Angadi , M. M. Kodabagi(2009) , "A Texture Based Methodology For Text Region Extraction From Low Resolution Natural Scene Images ", International Journal Of Image Processing (Ijip) Volume(3), Issue(5).

10. Strouthopoulos, N. Papamarkos, A.E. Atsalakis(2002)," Text Extraction in Complex Color Documents", Pattern Recognition 35 pp:1743–1758 .

11. Badreddine Khelifi, Nizar Zaghden, Adel M. Alimi, Rémy Mullot(2008), "Unsupervised Categorization Of Heterogeneous Text Images Based On Fractals", Proceedings of ICPR',pp.1-4.