ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 2, Issue 4, April 2016

Novel Text Recognition and Conversion in Real Time Scene Images

Jeraphin.J[1], Premalatha.J[2]

PG Scholar, Department of CSE, Vandayar Engineering College, Thanjavur, India [1]

Assistant Professor, Department of CSE, Vandayar Engineering College, Thanjavur, India [2]

***Abstract*—Detecting text in natural images, as opposed to scans of printed pages, faxes and business cards, is an important step for a number of Computer Vision applications, such as computerized aid for visually impaired and robotic navigation in urban environments. Retrieving texts in both indoor and outdoor environments provides contextual clues for a wide variety of vision tasks. In this project, we implement two processes such as text detection and text recognition. In text detection, use contrast map is then binaries by median filter and combined with edge map to identify the text stroke edge pixels based on feature extraction. The features are color, shape and texture features. Then implement text recognition. The first one is training a character recognizer to predict the category of a character in an image patch. The second one is training a binary character classifier for each character class to predict the existence of this category in an image patch. The two schemes are compatible with two promising applications related to scene text, which are text understanding and text retrieval.**

***Index Terms*— text detection, text recognition, feature extraction, text stroke.**

I. INTRODUCTION

Human machine system engineering is different from the more general and well known fields like human–computer interaction and socio-technical engineering in that it focuses on complex, dynamic control systems that often are partially automated (such as flying an airplane), it also studies human problem-solving in naturalistic settings or in high-fidelity simulation environments. Text characters and strings images can provide valuable information for many applications. The main objective is to detect and recognize text in images and convert into another format using text trajectory patterns so implement text detection and recognition approach. Using motion path analysis to identify text strings with trajectory patterns. Reading words in unconstrained images is a challenging problem of considerable practical interest. While text from scanned documents has served as the principal focus of Optical Character Recognition (OCR) applications in the past, text acquired in general settings (referred to as scene text) is becoming more prevalent with the proliferation of mobile imaging devices. Since text is a pervasive element in many environments, solving this problem has potential for significant impact. For example, reading scene text can play an important role in navigation for automobiles equipped with street-facing cameras in outdoor environments, and in assisting a blind person to navigate in certain indoor environments. In this benchmark, the organizers identified four sub problems: (1) cropped character classification, (2) full image text detection, (3) cropped word recognition, and (4) full image word recognition. The work of addressed the cropped character classification problem (1) and showed the relative effectiveness of using generic object recognition methods versus off-the-shelf OCR. The works introduced methods for text detection (2). The cropped word recognition problem (3) has also recently received attention. While progress has been made on the isolated components, there has been very little work on the full image word recognition problem (4); the only other work we are aware of that addresses the problem. Existing contributions are two-fold: (1) Evaluate the word detection and recognition performance of the two-step approach consisting of a state-of-the-art text detector and a leading OCR engine. (2) Construct a system rooted in modern object recognition techniques. We show that our object recognition-based pipelines perform significantly better than one using conventional [13] OCR and also show that, surprisingly, an object recognition-based pipeline achieves competitive performance without the need for an explicit text detection step. This result provides a significant simplification of the end-to- end pipeline and blurs the line between word recognition and the more common object recognition problems studied in computer vision.

II. RELATED WORK

This approach allows studying a large number of users with varying backgrounds in a large number of realistic contexts with their own devices (low internal validity due to a high variance but high external validity). This allowed analyzing the typing performance of users whose behavior would have been significantly altered in a very controlled setting. Because external factors cannot be ruled out and we have little control over the participants the study has a low internal validity as

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 2, Issue 4, April 2016

there been no possibility to control any contextual factors. The flip side is that the diversity of the environment provides a higher external validity than common lab studies. [5] Henze.N, Rukzio. E

Dictionary-based disambiguation (DBD) is a very popular solution for text entry on mobile phone keypads but suffers from two problems: 1. the resolution of encoding collision (two or more words sharing the same numeric key sequence) and 2. Entering out-of-vocabulary (OOV) words. In this paper, we present SHRIMP, a system and method that addresses these two problems by integrating DBD with camera based motion sensing that enables the user to express preference through a tilting or movement gesture. SHRIMP (Small Handheld Rapid Input with Motion and Prediction) runs on camera phones equipped with a standard 12-key keypad.[6] .Kristensson.P and Zhai.S

It present *Walk Type*, a system that uses a touch screen device's built-in accelerometer t increase text entry accuracy while the user is walking Taking inspiration from image stabilization techniques of cameras to remove motion blur, Walk Type [4] Goel.M, Findlatr.L, and Wobbrock.O compensates for imprecise input by incorporating multiple features computed from the accelerometer data: displacement, acceleration, and inference about the user's movement. Additionally, Walk Type uses tap location and the finger travel distance during taps to improve the user's text entry accuracy, features that increase overall accuracy regardless of movement. Previous work on adaptive text entry has largely focused on adjusting key-press probabilities based on language models.

III. PROPOSED DESIGN

Information retrieval from images has become an increasingly important research area in recent years. The rapid growth of digitized image collections is due to the widespread use of digital cameras and recorders combined with inexpensive disk storage technology. Textual information contained in images can provide one of the most useful keys for successful indexing and retrieval of information. Keyword searches for scene text of interest within images can provide additional capabilities to the search engines. Most existing algorithms for text detection were developed to process binary document images and do not perform well on the more complex images. In past years, many different methods have been developed for text detection in color document images by taking advantage of document characteristics. For example, simple edge-based detection filters such as the edge detector have been proposed to detect text based on the fact that the text is brighter than the image background. Some methods also make an assumption that the text and background in a local region have relatively uniform gray levels so that the contrast information can be used to extract text. Most recently motion path have been offered as an alternative method for detecting text in images. However, training networks and adjusting parameters increase the complexity of the implementation. As a result, building complete systems for these scenarios requires us to invent representations that account for all of these types of variations. Indeed, significant effort has gone into creating such systems, with top performers integrating dozens of cleverly combined features and processing stages. Another potential strength of these approaches is that we can easily generate large numbers of features that enable higher performance to be achieved by classification algorithms. In this project, we'll apply one such feature learning system to determine to what extent these algorithms may be useful in scene text detection and character recognition.

A. *Image Acquisition*

Indexing images requires information about their content. This content is often strongly related to the textual information appearing in them, which can be divided into two groups: Text appearing accidentally in an image that usually does not represent anything important related to the content of the image. Such texts are referred to as scene text. Text produced separately from the image is in general a very good key to understand the image. It is called artificial text. In this module, get images from users through web cameras, captured images, mobile images and so on.

B. *Gray Scale Conversion*

RGB image is converted into gray scale image. And eliminate noises from images using median filter and apply image binarization technique to separate background from foreground. The RGB value of a color is (r, g, b), where r, g, and b are integers between 0 and 255. The grayscale weighted average, x, is given by the formula

$$x = 0.299r + 0.587g + 0.114b.$$

C. *Median filter*

Then remove the noises from images by using filter techniques. The goal of the filter is to filter out noise that has corrupted image. It is based on a statistical approach. Typical filters are designed for a desired frequency response. Filtering is a nonlinear operation often used in image processing to reduce "salt and pepper" noise. A median filter is more effective than convolution when the goal is to simultaneously reduce noise and preserve edges. And implement image binarization tasks. Document Image Binarization is performed in the preprocessing stage for document analysis and it aims to segment the foreground text from the document

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 2, Issue 4, April 2016

background. A fast and accurate document image binarization technique is important for the ensuing document image processing tasks.

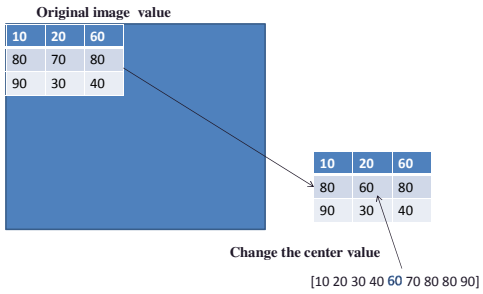

Fig 1 Median Filter

D. *Text Detection*

Analyze text strokes using feature extraction algorithm such as motion detection. In computer vision and image processing the concept of feature detection refers to methods that aim at computing abstractions of image information and making local decisions at every image point whether there is an image feature of a given type at that point or not. The resulting features will be subsets of the image domain, often in the form of isolated points, continuous curves or connected regions. Feature detection a low-level image processing operation. That is, it is usually performed as the first operation on an image, and examines every pixel to see if there is a feature present at that pixel. If this is part of a larger algorithm, then the algorithm will typically only examine the image in the region of the features. As a built-in pre-requisite to feature detection, the input image is usually smoothed by Gaussian kernel in scale-space representation and one or several feature images are computed, often expressed in terms of local derivative operations.

E. *Text Recognition*

After detecting text region in the image, from that text region text is extracted from the image using character descriptors and structure configuration. Segmentation is a process of distinguishing lines, words, and even characters of a hand written or machine printed document, a crucial step as it extracts the meaningful regions for analysis. There exist many sophisticated approaches for segmenting the region of interest. For handwritten document, this is quiet difficult. The details of line, word and character segmentation are discussed as follows.

1. Line Segmentation

Obviously the ascenders and descanters frequently intersect up and down of the adjacent lines, while the lines of text might itself flutter up and down. Each word of the line resides on the imaginary line that people use to assume while writing and a method has been formulated based on this notion. The local minima points are calibrated from each component to approximate this imaginary baseline. To guesstimate and categorize the minima of all components and to recognize different handwritten lines clustering techniques are deployed.

2. Word and Character Segmentation

The process of word segmentation succeeds the line separation task. Most of the word segmentation issues usually concentrate on discerning the gaps between the characters to distinguish the words from one another other. This process of discriminating words emerged from the notion that the spaces between words are usually larger than the spaces between the characters. There are not many approaches to word segmentation issues dealt in the literature. In spite of all these perceived conceptions, exemptions are quiet common due to flourishes in writing styles with leading and trailing ligatures. Alternative methods not depending on the one-dimensional distance between components, incorporates cues that humans use. Meticulous examination of the variation of spacing between the adjacent characters as a function of the corresponding characters themselves helps reveal the writing style of the author, in terms of spacing. The segmentation scheme comprises the notion of expecting greater spaces between characters with leading and trailing ligatures. These methods used to convert images with text into editable formats and processes input images with text and get editable documents like TXT file. Accuracy rate (AR) and false positive rate (FPR), are calculated to evaluate the performance of queried character classification. FPR represents the ratio between the number of incorrectly predicted negative samples and the total number of negative samples.

IV. System Architecture

The overall system design is the conceptual model that defines structure, behavior and more views of a system. An architecture description is a formal description and representation of a system.

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 2, Issue 4, April 2016

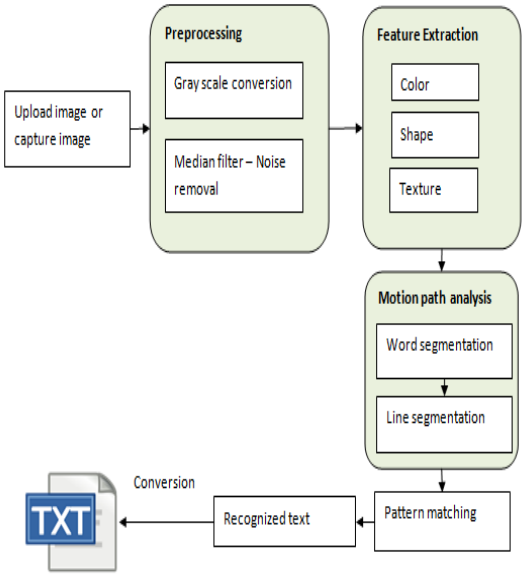

Fig 2 Block Diagram

V. EXPERIMENTAL ANALYSIS

Get the image from web camera or browse image from user.

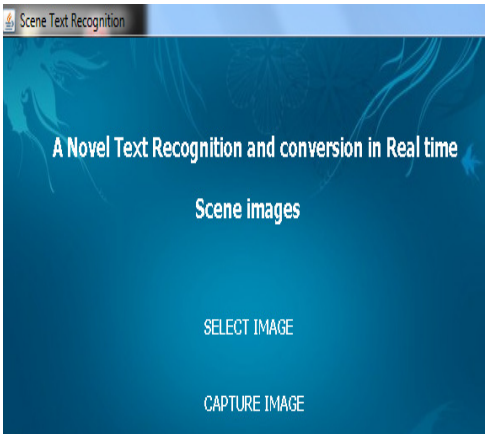

Fig 3 Scene Text Recognition

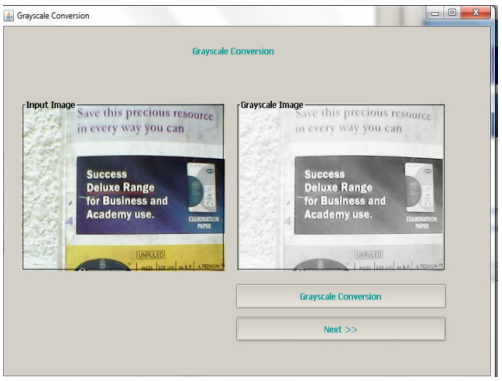

Fig 4 Gray Scale Conversion

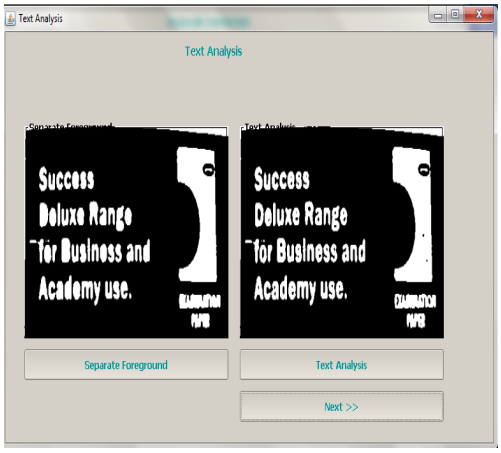

Fig 5 Text Analysis

Find text strokes from images based on image features extraction algorithms. Feature extraction is used to extract color, shape and texture features used to detect text in images.

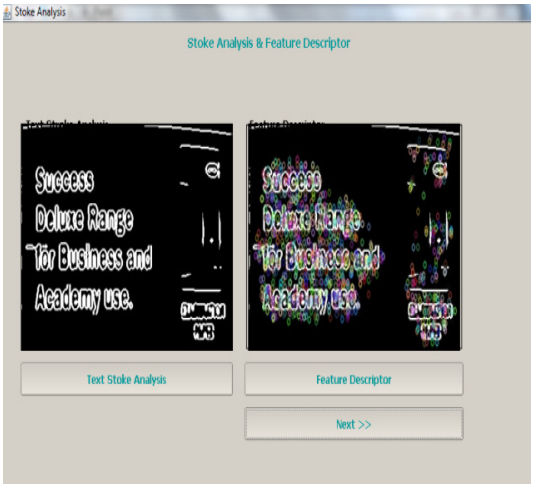

Fig 6 Store Analysis & Feature Descriptor

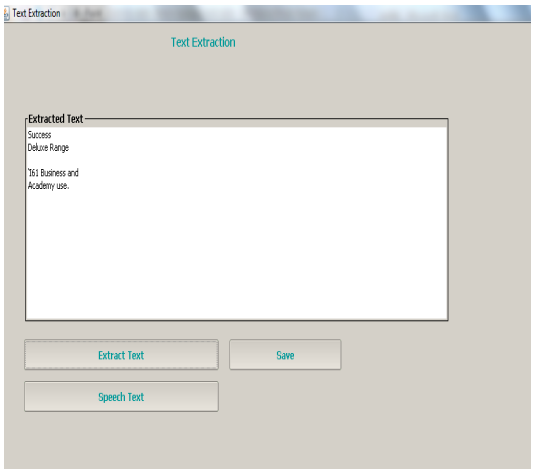

Fig 7 Text Extraction

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 2, Issue 4, April 2016

VI. CONCLUSION

The detects text regions from natural scene image and recognizes text information from the detected text regions. Text recognition method using multiple images, it employ motion path analysis model to configure text words and identify text strings with trajectory patterns where the output score of feature extraction is used as node features and the frequency of neighboring features are used as edge features. Text characters and strings images can provide valuable information for applications. Improve accuracy rate for text recognition in scene text application. To detect and recognize text in images and convert into another format in text trajectory patterns it improve the accuracy rate of text detection, and add lexicon analysis to extend our system to word-level recognition. To improve the accuracy and practicality of scene text extraction, the design more representative and discriminative features to model text structure.

REFERENCES

[1] Agrawal.S, Constandache.I , Ganokar.S and Caves.K,(2011),'Using mobile phones to write in air,' in Proc.9th Int.Conf.Mobile Syst.,Appl.Services,pp. 15-28.

[2] Amma.C,Georgi.M,and Tanja.S, (2012),'Airwriting: Hands-freee mobile text input by spotting and continuous recognition of 3D-space handwriting with inertial sensors,' in Proc. Int.Symp.Wearable Comput.,pp,52-59.

[3] Gomez.I, Anaya.P, Cabrera.R, and Molina.A ,(2012), 'Predictive system text entry controlled by accelerometer with any body part, 'J.Accessibility Desig,VOL.2,PP.31-44.

[4] Goel.M,Findlatr.L,and Wobbrock.O,(2012), 'Walktype: Using the accelerometer data to accommodate situational impairments in mobile touch screen text entry' in Proc.SIGCHI Conf. Human Factor Comput.Syst. 2687-2969.

[5] Henze.N and Rukzio. E, (2012) 'Observational and experimental investigation of typing behaviour using virtual keyboards for mobile devices,' in Proc. SIGCHI Conf. Human Factors Comput. Syst., ,pp. 2659–2668.

[6] Kristensson.P and Zhai.S, (2006),' SHARK2: A large vocabulary shorthand writing system for pen-based computers,' in Proc. 17th Annu. ACMSymp. User Interface Softw. Technol., pp. 43–52.

[7] Mac Kenzie.I.S,Castellucci.S.J,(2012) 'Reducing visual demand for gestural text input on touchscreen mobile devices,' in Proc Extended Abstracts Human Factors Comput. Syst., ,pp. 2585-2590.

[8] Tinwala.H and Mac Kenzie.I.S,(2010),' Eyes-free text entry with error correction on a touch screen mobile devices.' In Proc .6th Nordic Conf. Human- Comput. Interaction:Extending Boundaries,pp.511-520.

[9] Swiftkey, (2010). "Swiftkey app," [Online]. Available: http://www. Swiftkey.net.

[10] VisionObjects. (2013). Myscript. [Online]. Available: http://www. Visionobjects.com/ myscript.

[11] Diotek. (2013). Diopen. [Online]. Available: https://play.google.com/store/apps/details id=com.diotek.ime.diopen

[12] PhatWare. (2013). Writepad. [Online]. Available: http://www.phatware.com/q=product/details/writepad.

[13] Sony, NSZ-GS7. (2013) . [Online]. Available: http://www.sony.com.uk/product/google-tv/nsz-gs7.