Texture-based Real-Time Character Extraction and Recognition in Natural Images

Arun Kumar Singh

Abstract: Extraction and Recognition of text from any natural unstructured image is challenging. In this paper, an algorithm has been proposed which provides an efficient extraction of the text from the images for translation purposes. Optical character recognition (OCR) has been used to recognize the extracted information from images with the help of Prewitt operator. Pre-processing of images is carried out to enhance the desired information. Whereas noise and artefacts are removed from image using various filters to get optimized and quick results. From the experimental results, it can be concluded that the developed OCR system shows 100% accurate results in 5 scenarios out of 9 whilein the other cases the obtained results are near to desired results.

Index Terms: OCR, Detection, Translation, Prewitt operator, Noise and Artefact removal, Recognition.

I. INTRODUCTION

OCR is a unique tool used to recognize text from any image. It is basically developed by using the neural network and machine learning concepts, in which neural network is trained by making a database of different kind of images of same characters. Further, machine learning helps in making the database or the retrieving information from it for example k-nearest approach etc. There are many complex and tough languages in the world that are hard to understand or translate. To overcome these problems many techniques have been proposed which can help in translating the languages just by pointing the camera on it, for example, google glasses etc. In continuation of the research in this field, in the present work, a real-time character extraction and recognition algorithm using the texture based approach has been proposed to extract and recognize text or character from natural images. Most conventional algorithms uses canny edge detector to find out the edges for detection of characters [4] but in this proposed algorithm convolution of Prewitt operator and image is used to calculate the edges of an image. In the remainder of the paper, section 2 contains the related work. Section 3 and section 4 give the detailed explanation about the designed problem and proposed methodology, respectively. Further, a discussion of the experimental results is carried in section 5. Finally, section 6 contains the concluding remarks and scope for future work.

II. RELATED WORK

Since the beginning of 21st century, many algorithms are proposed to detect and recognize text from natural images in order to translate complex foreign languages into a simple language so as to provide convenience to the user. Many popular and existing technologies are working over this, for example, Fuji Xerox etc.The Maximally Stable Extremal Regions (MSER) feature detector works well for finding text regions [1]. It works well for text because the consistent color and high contrast of text lead to stable intensity profiles. Although The MSER algorithm picks out most of the text, it also detects many other stable regions in the image that are not text. A rule-based approach is used to remove non-textual regions. For example, geo- metric properties of text can be used to filter out non-textual regions using simple thresholds. Alternatively, a machine learning approach is deployed to train a text vs. non-text classifier. Typically, a combination of the two approaches produces better results [4]. This proposed approach uses a simple rule-based approach to filter non-text regions based on geometric properties. Another common metric used to discriminate between text and non-text is stroke width [3]. Stroke width is a measure of the width of the curves and lines that make up a character. Text regions tend to have little stroke width variation, whereas non-text regions tend to have larger variations. Currently, three basic approaches are being used for detection of text and these approaches can be categorized as:

A. Texture-based Approach

The object detection with sliding windows is common among these approaches. Different textural properties of text are used to extract candidate sub-windows and these sub-windows are merged finally to evaluate the final output. Following approaches can easily deal with artifacts in the background scenes. But some unwanted regions are always present in the outputs which consume a large amount of space and time.

B. Region-based approach

The region-based approach mainly focuses on similarity standards of text, such as size, color, stroke width, edge and gradient information, pixels are together into connected components (CCs) and non-text CCs are filtered out with a geometric hypothesis orconditional random fields (CRFs). These approaches usually have lowers the computation complexities and the outputs can closely cover the text regions. However, their algorithms face more challenges in litter background.

C. Hybrid Approach

Hybrid approaches seek to introduce a textural property of text regions into the region-based approach.

Revised Manuscript Received on May 19, 2019

Arun Kumar Singh, Asst. Professor, College of Computing and Informatics, Saudi Electronic University, Kingdom of Saudi Arabia-KSA

These approaches take advantages of both region-based approaches which can closely cover text regions and texture-based approaches which can estimate the course text location in litter scenes. However, they are also time-consuming.

III. PROBLEM STATEMENT

The main aim is to develop an algorithm which automatically reads text from every kind of natural image and extract the information or text for translation under certain circumstances.Mainly an image is classified into two types:

a. Structured scenes.

b. Unstructured scenes.

Currently, many techniques are being utilized for structured scenes. In which the scenes, as well as the text location in the scene, is predetermined. The present work focuses on the regions of an image that contains text. It is performed in unstructured scenarios. These scenarios contain undetermined or random scenarios. For example, to detect and recognize text automatically from captured video to alert a driver about a road sign. This is different than structured scenes.

Generally, during the acquisition of images, some difficulties are faced such as:

a. Presence of noise and artifacts in the captured images (light condition, reflection, blurring, shadows etc.).

b. Variation in the written text such as different font styles and font sizes.

c. Variation in devices that are used for capturing the images (VGA camera, low megapixel camera phone).

These difficulties are overcome in this proposed work by converting them to binary images without losing the region of interest. Binary images are not affected by light conditions. Firstly, the text is localized in the natural image then the conversion takes place [2]. Noise and artifacts such as granular noise are removed by the use of filters.

IV. METHODOLOGY

The proposed algorithm is mainly divided into four major sections. Firstly, pre-processing of the images is carried out to remove physical anomalies after which filter and operators are applied for detection of edges. Further, noise and artifacts are removed so that extraction and recognition of character can be done.

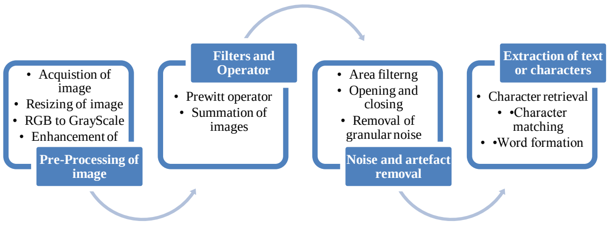


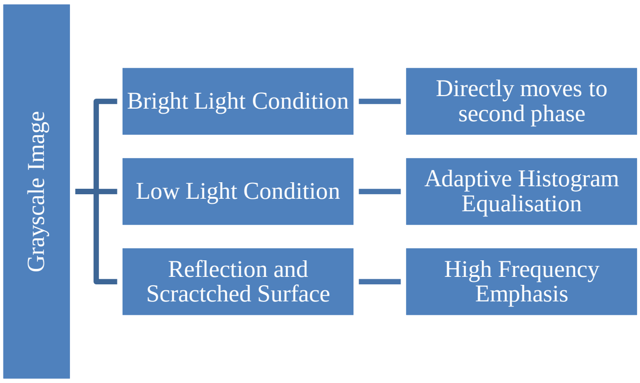
Fig. 1 Block diagram of the proposed algorithm

A. Pre-Processing of Image

Natural images are divided into three categories for pre-processing first is bright light condition images (BLC) second is low light condition images (LLC) and the last is reflective and scratched surfaces images (RSS). During pre-processing, the image is first, acquired and resized, then converted into two-dimensional matrix data from three-dimensional matrix data. The image acquired is in the form of a matrix of the pixel that is needed to be adjusted or resized as per requirement. It is necessary because the image acquired may be from different means like mobile phones, low sized screenshot, VGA and web camera which produce different size of images and may take very long time to acquire the text area or it will not produce the appropriate output. There is a need to increase or decrease the size hence providing a platform where computation is reduced to an optimum level. This task is done by resizing the captured image. After resizing the image so obtained needs to be converted into a gray image. The conversion of RGB image is too narrow down the matrix elements range and reduce down the computations. As the RGB image obtained is a three-dimensional matrix. Hence it has high complexity in manipulating or processing it. An image is consisting of three major components, brightness (intensity), hue and saturation. Grayscale images are composed of shades of gray, varying from black to white (0 to 255 levels). Whereas, an RGB image contains the mix of three colors, and require all the three properties to assimilate an image. So, there is a need to convert an RGB image into the gray scale image to eliminating the hue and saturation component from the former image. Now after the conversion, enhancement of the image is necessary to extract out the desired information from the image shown in (see fig. 3.a). So the image is categorized into three bright light conditions are directly moved to next phase of application of filter and operators. Low light image's contrast is enhanced with the help of adaptive histogram equalization (AHE). And in last reflective and scratched images are enhanced with the help of High-frequency emphasis function (HFE) [6] as the main objective is to get a final image from which text extraction is easier compared to the captured image.

$$H_e(u, v) = a + b\, H_{hp}(u, v) \tag{1}$$

Where a and b will be positive and a value will lie in the range of 0.25 to 0.5 and b>a therefore its value will lie in the range of 1.5 to 2.5.



Grayscale Image
Bright Light Condition
Directly moves to second phase
Low Light Condition
Adaptive Histogram Equalisation
Reflection and Scractched Surface
High Frequency Emphasis

Fig. 2 Enhancement of image after conversion of captured image into grayscale image

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

B. Application of Filters and Operators

Afterpre-processingoftheimagetheedgesofthegrayimageareevaluateditwill differentiatebetweenthenoise,artefactsandareaofinterest.Theedgedetectionisdone withthehelpofPrewittoperator.Thisoperatorisusedinimageprocessing,majorly withedgedetectionalgorithms.ThePrewittoperatorisusedforconvolutingtheimage withasmall,separable,andintegervaluedfiltermatrix,inhorizontalandverticaldirections.Theoperatorcalculatesthegradientoftheimageintensityateachpoint,givingthedirectionofthelargestpossibleincreasefromlighttodarkandtherateofchange inthatdirection.Theresultshowsthechangesinthepixelvaluesanddirectionwhich aremagnifiedaftertheapplicationoffiltersandoperators.

C. Prewitt Operator

ThesearebasicallyofthreetypesofPrewittoperator[7]andtheoperatorusesthree 3×3kernelswhichareconvolvedwiththeoriginalimagetocalculateapproximations ofthederivatives–horizontal,verticalanddiagonalchange.Aisdefinedasthesource image,andGxandGyaretwoimageswhichateachpointcontainthehorizontaland verticalderivativeapproximations,whichcanbecomputedas:

$$Gx = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} * A \qquad (2)$$

$$Gy = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} * A \qquad (3)$$

Where “*” here denotes the 1-dimensional convolution operation.

The x-coordinate is defined here as increasing in the right direction, and the y-coordinate is defined as increasing in the down direction. At each point in the image, the resulting gradient approximations can be combined to give the gradient magnitude, using:

$$G = \sqrt{Gx^2 + Gy^2} \qquad (4)$$

Gradient’s direction is calculated by

$$\Theta = \text{atan}^2 (G_x ,G_y) \qquad (5)$$

ϴ is 0 for a vertical edge which is darker on the right side.

D. Noise and Artefacts Removal

A histogram is developed after the application of filters and operator which is used to analyze the values of the pixel present in the image ‘I’. By analyzing (fig. 3.b) it is observed that most of the pixels present are toward zero only which means that portion that it has more dark part than the white part. As observed in the (fig. 3a) the text has all white portion and its value is towards 256. Afterwards, the thresholding is carried out to differentiate pixels in two category white (high value) and black (low value). There is an advantage that the obtained image consists of only two-pixel values that reduce down the computation time significantly.

An area filter is used to remove the other entire component and keeps only the required region of interest. A specific range of areas is allowed under which the filter allows only connected component of that area range rest other pixels are rejected. Morphological operations are applied after area filtering such as opening by using diamond structuring element. After this process the output comes out as shown inn (fig. 3.d).

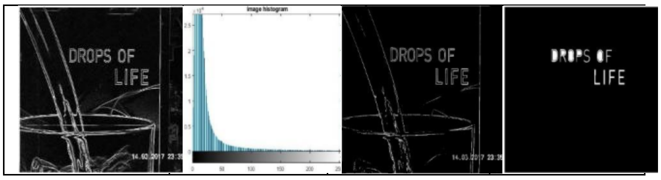
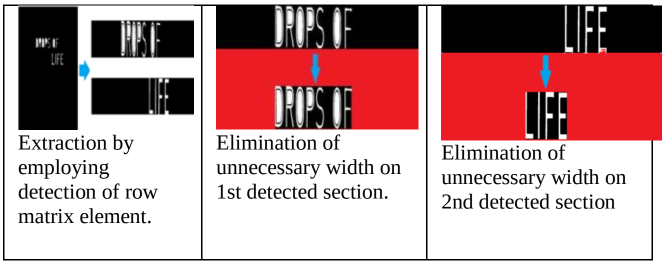


Fig. 3 (a) Resultant image “I” (b) histogram of resultant image “I” (c) Binary image obtained after thresholding (d) image obtained after noise and artefacts removal

E. Extraction of Text or Characters

Afterward,croppingoftheimageisdonetoextractoutthetextand characters,but itisanunstructuredscenario,therefore,locationisunknownsolocalizationoftext takesplacebyextractingdifferenttexturalpropertiesoftextwhichareusedtoextract candidatesub-windowsandthesesub-windowsaremergedfinallytoevaluatethefinal output[4].Firsthorizontalsub-windowsareusedtodetectlinesfromimagethenverticalsub-windowsareusedtodetectcharacterfromtheimageasshownin(seefig.4). Withthis,everypartgetextractedwithsomeextracomponentsi.e. thespacebetween twocharactersandithasallvalueszeroinitjustlikethecharacterspaceisverymuch smallerinsize.Thereforematrixwithallelementszeroandlessinsizeisremoved,thus protectingwordsspaceasshownin(fig.5aand5b).

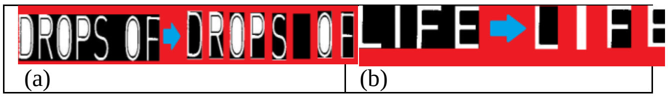


Extraction by employing detection of row matrix element.

Elimination of unnecessary width on 1st detected section.

Elimination of unnecessary width on 2nd detected section

Fig. 4 Extraction of words from different lines



(a) (b)

Fig. 5 Extraction of characters from words (a) Words segmentation for 1st section (b) Words segmentation for 2nd section.

V. EXPERIMENTAL RESULT

Atechniquehasbeendevelopedfordetectionaswellasextraction oftextandcharacter.Robustclassesandfunctionsofimageproces

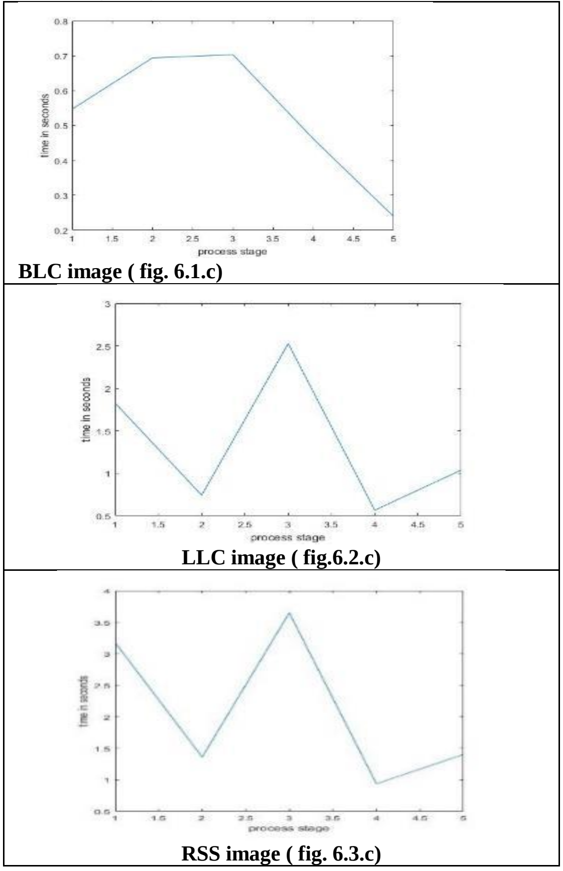
Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

singtechniquesareemployedinthis proposedalgorithm.Capturedimageisconvertedintobinaryimagethenrow-column detectionisdoneafterthatlettersareextractedfromword.Extractedlettersarematched withthetemplatestorecognizethecharacter.Shownin(fig.6)Nowthetimeisdifferentfordifferentcategoryofimagesasitisobservedfromtable2tisclearfromthe timecurvethatnoiseandartefactsremovaltookthemosttimeoftheproposedsystem.

Table 1 Detailed Analysis of different Natural Images

Image Type	**Original Image size**	**Resize Factor**	**No. of edges**	**Total time (in sec)**	**AHE**	**HFE**
Fig. 6.1.c	**4.11MB 3456x4608**	**0.3**	**131392**	**10.5301**	**No**	**No**
Fig. 6.2.c	**3.90 MB 4608x3456**	**0.1**	**136383**	**7.4773**	**Yes**	**No**
Fig. 6.3.c	**1.71 MB 1605x2185**	**0.1**	**117186**	**2.6444**	**No**	**Yes**

Table 2 Graphical Representation of time taken in corresponding stages



BLC image (fig. 6.1.c)

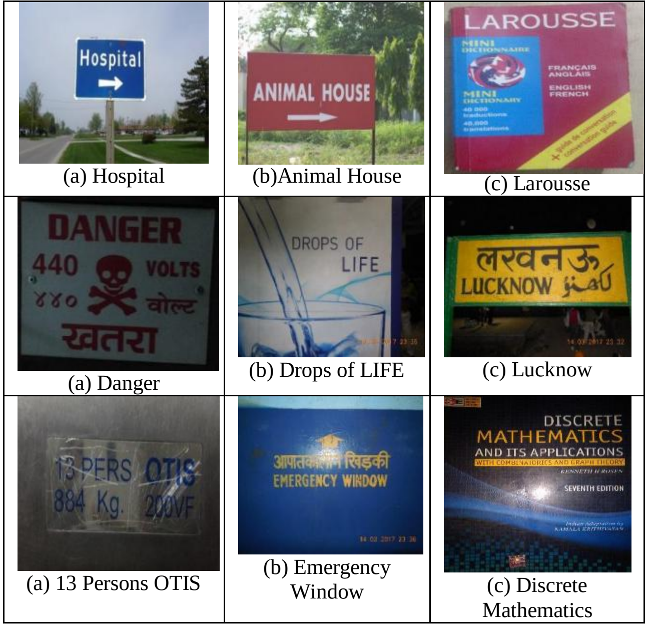
LLC image (fig.6.2.c)

RSS image (fig. 6.3.c)

Table 3 Accuracy and Error Analysis of Proposed System

Image Type	Image No.	Total Characters	Correctly matched character	Mismatched Character	Missed Character	Accuracy (in %)	Error (in %)
BLC Image	Fig (6.1.a)	8	8	0	0	100	0
	Fig (6.1.b)	11	11	0	0	100	0
	Fig (6.1.c)	16	8	0	8	50	0
LLC Image	Fig (6.2.a)	14	11	3	0	78.57	27.27
	Fig (6.2.b)	11	11	0	0	100	0
	Fig (6.2.c)	7	7	4	0	100	57.14
RSS Image	Fig (6.3.a)	20	14	2	4	70	10
	Fig (6.3.b)	15	15	0	0	100	0
	Fig (6.3.c)	63	50	0	13	79.36	0

Accuracy and error analysis of every image used in this proposed system is done in(table 3) in this three parameters are calculated to estimate accuracy and error ofproposed system. Matched character parameter is used to calculate accuracy and mismatched character parameter is used to calculate error in the system.



(a) Hospital (b)Animal House (c) Larousse

(a) Danger (b) Drops of LIFE (c) Lucknow

(a) 13 Persons OTIS (b) Emergency Window (c) Discrete Mathematics

Fig. 6 (a) BLC Image (b) LLC Image (c) RSS Image

VI. CONCLUSION

Recognizing and extraction of character from natural images is a challenging task, till dietitian area of research to make the machine capable of detecting characters with hundred percent efficiency from real-time images. Edge detection algorithm tool isalsoused, called Prewitt operator to detecttheedgespresentinourimage. Otherfilterslikebinaryimagepropertyfilter,areafiltersareusedtoassesstheimage forcertaintypeofpatterns,whicharecalledattributeoftheimagelike,eccentricity, solidity,orotherEulernumberconstraints.Alltheseattributesimposedrestrictionson theinputimagetopasstheseconditions.Thiswillomittheunwanteddataandenhance thedataanticipated. Thisresearchfieldhasnocloseendsandhasimmenseapplications,takingintoconcernsthefutureprospectsitcanbeclearlyseenthatcombiningthisconceptwith MATLABwindowsapplicationintegrationandgraphicaluserInterface,imageprocessingtoolsofMATLABenvironmentcanbeeasilydesigned.Asitisobservedfrom (seetable1andtable2)thatthetimerequiredforbigimagesishight omakearealtime systemoutputtimeshouldbereducedandfigure6.2.a,6.3.band6. 2.ccontainHindi languagebesidesEnglishsothefuturesystemcantranslatemultip lelanguagesata time.

REFERENCES

1. Chen Huizhong, Sam.S.Tsai, Georg Schroth, David M.ChenRadekGrzeszczuk, Bernd Girod. "Robust Text Detection in Natural Images with Edge-Enhanced Maximally Stable Extremal Regions." Image Processing (ICIP), 2011 18th IEEE Inter- national Conference on. IEEE, 2011. doi:10.1109/icip.2011.6116200
2. Gonzalez Alvaro, Luis M. Bergara, J Javier Yebes, Sebastian Bronte. "Text location in complex images." Pattern Recognition (ICPR), 2012 21st International Conference on. IEEE, 2012.
3. Li, Yao, and Huchuan Lu. "Scene text detection via stroke width." Pattern Recognition (ICPR), 2012 21st International Conference on. IEEE, 2012.
4. Neumann, Lukas, and Jiri Matas. "Real-time scene text localization and recognition." Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012. doi:10.1109/CVPR.2012.6248097.
5. Gonzalez, Rafael, C; Woods, Richard E (2008). Digital Image Processing, 3rd Edition. Pearson Prentice Hall. p. 577,188.
6. Philip Howard. "True-Color GIF Example". Archived from the original on 22 February 2015. Retrieved 23 March 2012.
7. J.M.S. Prewitt "Object Enhancement and Extraction" in "Picture processing and Psychopictorics", Academic Press,1970
8. Michael Freeman (2005). The Digital SLR Handbook. Ilex.
9. Stephen Rawls, Hauigo Cao, EkraamSabir and Prem Natarajan "Combining deep learning and language modeling for segmentation-free OCR from raw pixels" IEEE International conference on Arabic script Recognition and Analysis (ASAR), 2017. doi: 10.1109/ASAR.2017.8067772
10. Guillaume Chiron, Antoine Doucet, MickaelCoustaty, Muriel Visani, Jean- Philippe Moreux. "Impact of OCR Errors on the Use of Digital Libraries: Towars a better Access to information" ACM/IEEE Joint Conference on Digital Libraries (JCDL), 2017. doi: 10.1109/JCDL.2017.7991582
11. Mohammad AzimUlEkram, AnjaniChaudhary, AshutoshYadav, JagadishKha- nal, SemithAsian."Book organization checking algorithm using image segmenta- tion and OCR".IEEE 60th International Midwest Symposium on Circuits and Sys-tems (MWSCAS), 2017. doi: 10.1109/MWSCAS.2017.8052894

AUTHORS PROFILE

Dr. Arun Kumar Singh is working as an Asst. Professor in the College of Computing and Informatics (Saudi Electronic University, KSA). He received Ph.D. in CS/IT under the guidance of Dr. Neelam Srivastava (IET Lucknow) and Dr. R. P. Agarwal (IIT Roorkee), M.Tech. (IT-WCC) degree from IIIT-Allahabad under the guidance of Prof. M. Radha Krishnan and B.E. (ECE) in 2002 from Dr. B.R. Ambedkar University, Agra, UP, India. His research interests are IoT, HCI, Big Data, Network Management, Wireless networking, Social Networking and Mobile computing.