Scientific Journal of Impact Factor(SJIF): 3.134
e-ISSN(O): 2348-4470
p-ISSN(P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 2,Issue 5, May -2015

A Novel Approach to Detect and localize the Text inNatural scene image

Prof. B. K. Sarojini, Mr. Mahesh Mahadar

Department of Electronics and Communication Engineering, Basaveshwara Engineering College, Bagalkot India

Abstract: *The informationavailablein the naturalscene provides very importantcluesformanyimagebased application. So the detectionandlocalization of textfromnaturalscene imageisimportant taskforcontent basedimageanalysis. In thispaper,wedevelop anovelapproach todetectand localizetextsin naturalsceneimages. Thisproblem ischallengingduetothe fact that text has different in size, alignment, orientation, style,complexbackground of imagesaswellasimage having the lowcontrast. Initially,theRGBimageisconvertedintograyscaleimage andapplied to thelocal binarizationalgorithmfor the segmentation. Afterthat,HaarWaveletTransform(DWT)isused which isfastestas compared withallwaveletsbecauseitsfilter coefficientsareeither 1 or -1. It decomposeimageintofoursub band, oneisaverage andotherthreearedetailedwhichhelps to find outtheapproximatelyconfidentregion.To filter out the non-text component, aconditional randomfield(CRF) is used which appliedonconfidentregionimage. Finally,by usingsome predefinedcondition,thetextis obtainedin boundingbox.*

Keywords:*Wavelet transform, Text Detection, Text Localization, Conditional random field (CRF).*

I. INTRODUCTION

Recently the digital image capturing devices,suchasdigitalcameras,mobilephones are increased to the large extent socontent based imageanalysistechniquesarereceiving intensive attentioninrecentyears.As compared with the other contentsin images,textinformationhasgreatadvantage becauseitiseasilyunderstood bybothhumanandcomputerandfinds wide range ofapplications but localization of text is very difficult task forthefollowingreasons.Firstofall, text has wide range of variety in font, orientation; style as well as size maychangesfromsmallto too large. Secondly,textspresentinanimage may have low contrast,multiplecolorsandappearin a muchcluttered background.[5]

Thetextdetectionandlocalization methodscan bemainlycategorized intotwogroups:connectedcomponent(CC) based and region based.Region based methods uses textureanalysis principle todetectandlocalizetextregions. It extracts thefeaturevectorfromeachlocalregion and fedintoaclassifier which estimatesthetext component. Thenneighboring textregionsaregroupedtogeneratetext blocks. As the textregionshas distincttexturalproperties fromnon-textones,thesemethodscan easilydetectandlocalizetexts evenwhenimagesarenoisy.Ontheotherhand, CC basedmethodsuses the edgedetectionorcolorclustering for thedirect segmentcandidatetextcomponents. Finally,thenon-textcomponents arethen removed with some specific rules but thismethodhaslowercomputationcostandthelocated textcomponentscanbeused as it isfor recognition. Still existing methods havesomeproblemsto be solved.Such as forregion based methods,thespeedis veryslow andtheperformance isalso sensitivetotextalignmentand orientation. For CC based methods it required a priorknowledgeoftextpositionandscale to segmenttext components accurately. [8]

II. RELATEDWORK

Most of theregion based methodsare mainlybasedonprinciple that textregionshavedifferentcharacteristicsfromnon-textregions suchasthedistributionofgradientstrengthandtextureproperties. Many efforts have been made for text extraction and recognition in image.

Weinman et al. provides amethod whichuses a Conditional Random Field (CRF) model for text detection. This model assigns the candidate components to one of the two classes such as “text” or “non-text” by considering the properties of unary components as well as relationship between the contextual components. This method provides the benefit over the traditional local region based text detection methods. [1]

Chung Wei Liang and Po Yueh Chen provide useful and effective approach to extract the text region from static image. They use Haar Discrete Wavelet Transform (DWT) to localize the text region along with the morphological operator to detect edges of candidate text region which is used for the isolation of text data from documented video image. [2]

The methodproposedby,Kim*etal.* uses the support vector machine (SVM) to analyze textural properties of text. The intensity of raw pixel which required for analysis of textual pattern are fed to the SVM then continuouslyadaptivemean-shift(CAMSHIFT)is applied to the result of texture analysis to identify the text region so the combination of SVM and CAMSHIFT provides a robust and efficient approach for text detection. [3]

X. L. Chen et al. provides the combined approach of multiscale edge detection and multiresolution, color analysis, adaptive searching as well as affine rectification in the proposed framework for sign detection, with different emphases at each phase to handle the text in different orientations, sizes, color distributions and backgrounds. It uses affine rectification to recover deformation of the text regions caused by an inappropriate camera view angle. The procedure can significantly improve text detection rate and optical character recognition (OCR) accuracy. [5]

Kim segments an image using color clustering in a color histogram of RGB space. Non-text components mainly contain the long horizontal lines as well as image boundaries these are eliminated by iterative projection profile analysis. To filter out the non-character components this method used cluster-based templates for multi-segment characters to lower down the difficulty in defining heuristics for filtering out non-text components. [4]

Zhang et al. uses the Markov random field (MRF) to detect the neighboring information of components. First mean shift algorithm is used to segment the candidate text components then after developing the component adjacency graph, first-order component term are integrated using MRF model and finally for labeling components as “text” or “non-text” a higher order contextual term is used. [7]

S.Audithan et al. uses the Haar discrete wavelet transform (DWT) which is the fastest among all wavelets because its filter coefficients are either 1 or -1. DWT detect edges and then based on the edge map, line feature vector graph is generated which helps to extract the stroke information. Finally text regions are generated and filtered according to obtain the line features. [10]

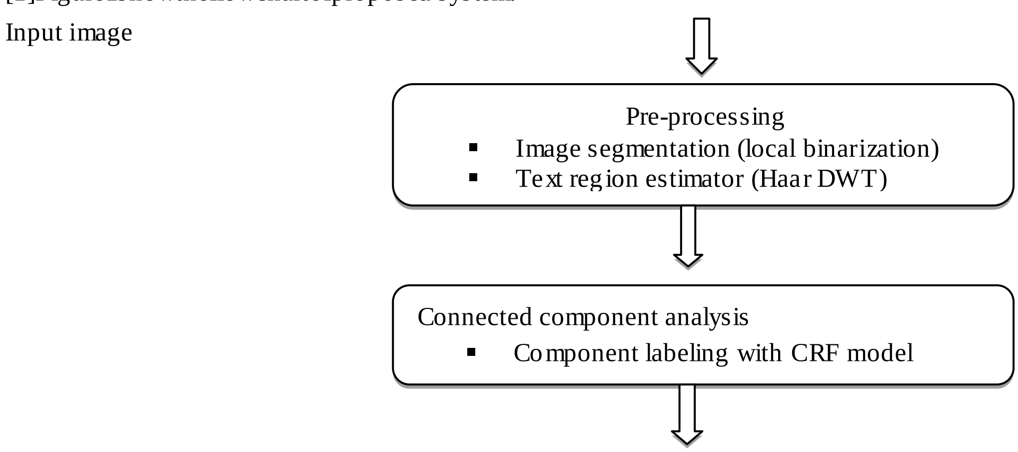
Zhu et al. uses the first nonlinear local binarization algorithm for the segmentation of candidate CCs. There are large numbers of component feature which contains geometry, edge contrast, shape regularity, spatial coherence features as well as stroke statistics are defined to train an AdaBoost classifier which helps to filter out non-text components.[6]

The approach of Khushbu C. Saneris mainly divided into three step 1)Pre-processing 2)Connected Component Analysis 3)Optical Character Recognition. In pre-processing step, the color image is converted into binarized image for the edge detection. A conditional random field (CRF) model use binary discourse part relationships and unary part properties along with the designed supervised parameter learning to filter out the non-text components. Then recognized text is localized in original image and text parts are classified into text lines. Once the line partition is over, character recognition will be done using Optical character recognition to recognize the character. Here the results are evaluated on the natural image dataset. [9]

Liu et al. provide the approach to detect color texts from natural scene images. It contains the combination of connected component based approach as well as region based approach. To detect the probabilities of text scale and position a text region detector is designed then an efficient local binarization algorithm is used to segment candidate text components. A conditional random field (CRF) model along with supervised parameter learning is designed to combine the binary contextual component relationships and unary component properties. Finally, learning-based energy minimization method is used to group the text components into text lines or words.[8]

III. SYSTEM OVERVIEW

To overcome the above difficulties, we present an approach to detect and localize texts in natural scene images with wavelet transform. A text region detector is designed using Haar discrete wavelet transform to estimate the probabilities of text position, and then segment candidate text components using an efficient local binarization algorithm. [6] For labeling the connected component as a text or non-text, we use the unary as well as binary properties of the conditional random field. Finally, text components are localizing with use of bounding box. [1] Figure 1 show the flowchart of proposed system.



Text grouping

- Bonding boxes (merging text)

⇩

Localization result

Figure1: Flow Chartof Proposed Model

VI. PRE PROCESSING

A text region detector is designed to estimate the text confidence and the corresponding scale which helps for efficiently utilize and extract local text region information, depends on which candidate text components can be segmented and analyzed accurately.

A. Image Segmentation

Initially RGB image is converted into gray-level image. Then Niblack's local binarization algorithm is used because of its high insensitivity and efficiency for degraded image.To segment candidate connected components (CCs) from the gray-level image, the following formula should be used to binarize each pixel which is defined as

$$m(x) = \begin{cases} 0, & \text{if gray } (x) < i(x) - k.\, s(x); \\ 255, & \text{if gray } (x) > i(x) + k.\, s(x); \\ 100, & \text{otherwise,} \end{cases}$$

Where s(x) and i(x) are the standard deviation(STD) and intensity mean of the pixels for a radius window which has pixel x is available at the center and the k is smoothing term which empirically set to 0.45. Most of the methods use the window having fixed radius or it is chosen based on some simple rules such as the gray-level standard deviation (STD) while in our method we use text scale map to calculate the radius of window which provides more stability under noisy conditions. In local binarization, we assume that within each local region, the foreground pixels has gray-level values must be higher or lower than the average intensity of the pixels,So the connected components with a value 0 or 255 are extracted as candidate text components while those of value 100 are non-text components so they are not considered further. [6]Figure 2 show the resultant image of image segmentation.

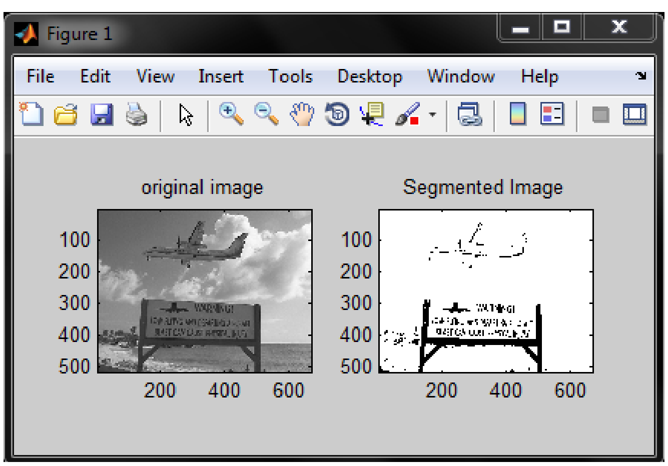


Figure 2: (a) original image (b) Segmented image

A. Text Region Estimator

If the input image is a gray-level image, such image is directly processedtoHaar discrete wavelet transform for text region estimation. Butif theinput imageiscolored,thenitsintensity image should be calculated by combing its RGB components.Normally,colorimages are capturedwith the help of digitalcameras.These pictures are mainly available in the Red Green Blue color space. Intensity imageYiscalculated as:

$$Y= 0.299R+ 0.587G+ 0.114B$$

These imageYis then finally processedwith2-DHaar discretewavelet transform. Here theYisactually represent the Valuecomponentofthe Hue Saturation Value (HSV) color space. So there is conversion fromRGBcolorspaceintoHSVcolorspace in abovestep, once it is over then the Valuecomponentisextractedfrom HSVcolorspaceusing above expression.To reduce the effect of noise in the imagemostly median filtering techniques are used which is applied on the abovegrayscaleimage. Afterthisfilteringstep, amajorpartof noisewillberemoved while theedgesintheimagearestill preserved. [2]

B. Haar discrete wavelettransforms

For multi resolutionrepresentation,a haardiscretewavelettransform is a very powerfultool for imageprocessing as well as signal analysis.Itcan decomposesignalinthe frequencydomainwith different frequencycomponents. Two dimensionaldiscretewavelettransformdecomposesaninput imageinto four components, one average component (LL) and three detail components(LH,HL,HH)bycalculationsoflow-passanda high-passfilter combination asshowninFigure3.

LL	HL
LH	HH

Figure3: The result of 2-D DWT decomposition

To detect the candidatetextedgesinthe originalimage the detailed sub-band component are used. Inimageprocessing,themulti resolutionof2-D DWT is mainly used to detect edges from the original image. The processingtime ofthe2-DDWTis much fasterthantraditional edge detectionfilters because it candetect threekindsofedgesatatime. But the result provided by2-D DWTcan besimilar as compared with the conventional edge detection filters.This is why wechooseHaarDWT becauseitisefficient and simplerthanthatofanyotherwavelets. [10]Figure 4show the approximate text area in image.

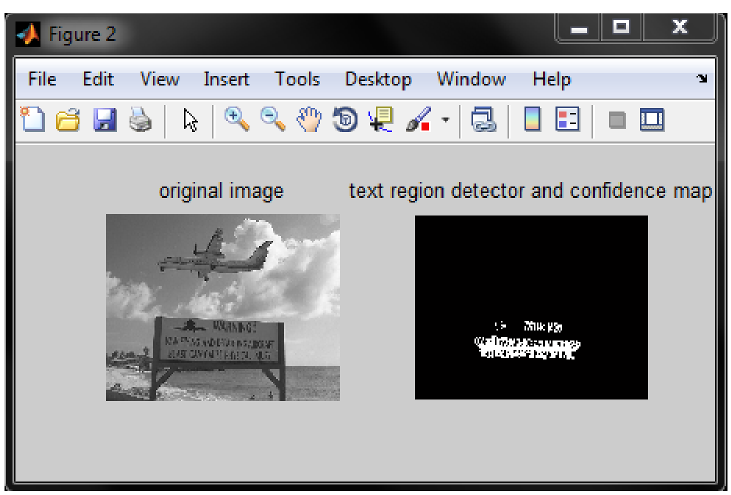


Figure 4: (a) original image (b) text region detector and Confidence map

V.CONNECTEDCOMPONENTANALYSIS

Forconnectedcomponents analysis(CCA),weassigneach candidatecomponentsto oneofthetwoclasses such as "text"and"non-text"byusing bothbinarycontextual componentrelationships and unarycomponentpropertiesof conditionalrandomfield(CRF)model.

A. *Introduction of CRF*

Conditionalrandom fields(CRFs)are mainlydesignedforlabelingtaskssuchasdocument image segmentation, text identification as well as naturallanguageprocessing.CRFs are basically probabilisticgraphicalmodeland used tolabelthetextregionfromdifferentareasinimages having the spatial interdependencies.Forexample,textblocksare sequentiallyavailable from left toright. So by considering theinformationofneighboring text blocks,isolatednoises available in these sequential textblocks can beeasilyremoved whichprovidemoreaccurate results of labeling. [1]So CRFs provide a more flexible formulationrather than the othergenerativegraphical models suchasMarkovrandom fields(MRFs)whichrequire specifying the likelihood function. [7]Moreformally,let assume the observedfeaturesfrom candidateblocks$X=\{a_i\}$,andrandom variablesovercorrespondinglabels$Y=\{b_i\}$.Thenjoint distribution overthelabelb_i*with* givenobservation$a_i$is represented as

$$d(b_i \,\mathrm{I}\, a_i)\ \alpha\ \exp\left(\lambda P(b_i, X) + \mu \sum_{(i,j)\,\varepsilon\,\mathrm{E}} R\,(b_i, b_j, \mathrm{X})\right) \quad (1)$$

Where thefunction$P(b_{\mathrm{i}},$ X) iscalledas associatedpotentialwhich measures the confidence of label $b_i$by considering theobservations, and function $R(b_{\mathrm{i}}$, b_{i}, X) is interaction potential which providesmoothlabelsoverentiregraph$G$,λandμparameters helps to control the influence from neighboringnodestocenter node$i$ and observations,and$(i,j)\,\varepsilon\,E$ represents the neighboringnodesofnode$R$whichareconnectedbyedges E inthegraph G.
Inourwork,weusethetopologyforourCRFs.Byconsidering aMarkovassumption,detectedblockintheimage areexclusivelyrepresentbyeachgray node $g_i$inthehiddenlayer andthen connectstoitsfour nearest neighbors blockalong withtheir corresponding observations. In the case ofreal images, the neighbor blocksare mainlydeterminedbyEuclideandistance betweenthembut itmaynotnecessarilybelocatedasa grid. Tointegratethepredictedconfidence of blocksintoCRFs framework, wedefine the associatedpotential as

$$P(b_i, X) = \sum_{j\varepsilon N} e_j\ \exp\left(-\left|s_{i,\,j}\,.\,\cos\,(\theta_{i,\,j})\right|\right) \quad (2)$$

Where$j$ runs over neighbors of node $i$includingitself, and $S_i$isthe spatialEuclideandistancebetweennode$i$,j and $e_j$is theposteriorwhich is estimatedbytheSVMfornode$i$,$j$andθisthe angle between centers of node$i$ and$j$. The idea behind equation2isthatiftwoneighboringnodesarevery closeto eachotheras well astheirseparationismostlyhorizontal,then theyhavemoreinfluenceson each other. [9]

B.PropertiesofCRF

a. UnaryComponent Features

Here we considerdifferent typesofunarycomponent featuresuchasAspect ratio,height, normalizedwidth and Compactness,Tocharacterizesinglecomponent's geometric andtextural properties. [8]

b.Binary Component Features

Here weconsiderdifferenttypes ofbinarycomponentfeatures suchasOverlapratio,Shape difference,Gray-level difference,Scaleratio, Tocharacterizethegeometricandspatialrelationshipand textural similarity between twoneighboring component. [8]

VI. TEXT GROUPING

Here adjacentlettersare grouped togethertoform words. For the performanceanalysisofa text extraction algorithm, it is recommended thattherecallrates andprecision must be compute. But theperformance parameters are mainly dependent on correctlyclassified words. Some proposed methods are effectivebut it is toocomplicated because oftraining data necessity while theothermethods are simplerbut not effectivefor text grouping. To overcome all such drawback weproposethe boundingboxes(BB) concept tomergeadjacent lettersinwordswhichis based onthe computationofdistancesbetween theseboxes(BB)of letters, detectedinthe abovestep.The parametersB1 and B2 are usedin themergingletterprocess which represents thecentercoordinates ofthetwoBBs of connectedcomponent. Figure 4 represent the merging process.

Here B1 (y1a), B1 (y2a)and B2 (y1a), B2(y2a) representthecoordinates ofthe firstandsecondBBsinverticaldirection respectively while Width 1 andWidth 2representthe widthofthefirst and second BBs respectively.'Distance'representsthedistancebetweenthecentroidsofthetwo BBs consideredin the horizontal direction.

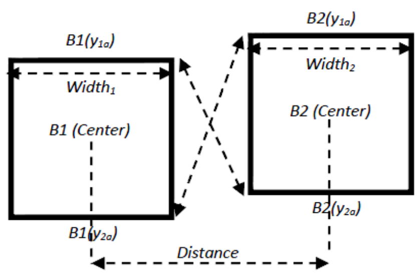


Figure4: Parametersusedin merging process

The first step for merging is basedonamerging ofletters along the horizontal line. Here we consider only those images which contain relatively well aligned letters. The conditions formerging thelettersinthe detected regionsaredefinedasfallows

- [B2 (y2a)>B1 (y1a)]&[B2 (y1a)<B1 (y2a)]
- [Distance< 0.7 ×Max (Width1, Width2)]

The pairofBBswhichsatisfyboth theaboveconditionsis thenmerged together inthisstep to obtain the word. [8] Figure 5showthe boundingboxon each textin input color image.

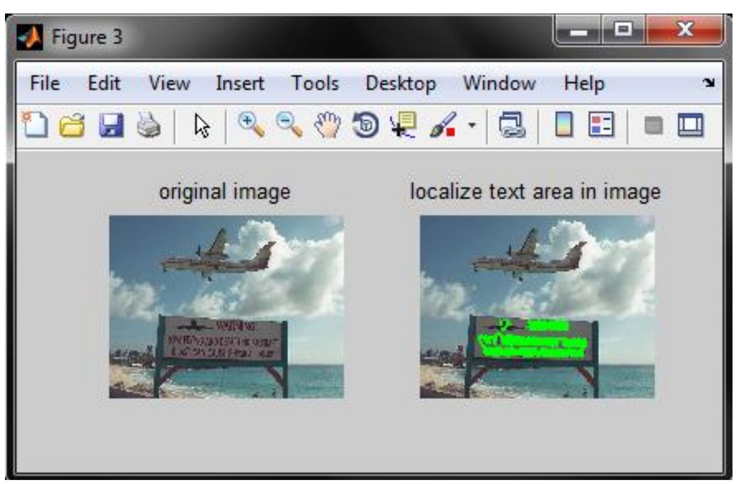


Figure 5: (a) Original image (b) localize text area in image

VII. CONCLUSION

Thegiven inputc o l o r imagestobeconvertintograyscaleimageand then image segmentation is carried out on thegrayscaleimage to obtain the segmented image. Wavelet Transformused to decomposethe segmentedimage.Itwilldecomposetheoriginalimageintofour frequency sub bands to improve the contrast and resolutionoftheimagewhich helps tofindthe approximated text area in theimageusingconnectedcomponent graph.Finally by usingboundingbox concept, we localizetext in image.So it providesrobust approach to detect and localize texts by integrating region information.

REFERENCES

1. J.Weinman,A.Hanson,andA.McCallum, "Signdetectioninnaturalimageswithconditionalrandomfields,"inProc.14thIEEEWork-shoponMachineLearningforSignalProcessing(MLSP'04),SãoLuis, Brazil,2004,pp.549-558.

2. Chung-Wei LiangandPo-YuehChen,"DWT based Text Localization", International Journal of Applied Science and Engineering: 2004

3. K.I.Kim,K.Jung,andJ.H.Kim,"Texture-basedapproachfortext detection inimagesusingsupportvectormachines andcontinuously adaptivemeanshiftalgorithm," IEEETrans.PatternAnal.Mach.In tell vol.25,no.12,pp.1631–1639,2003.

4. K.Jung,K.I.Kim,andA.K.Jain,"Textinformation extractioninimagesandvideo:Asurvey," Pattern Recogn.,vol.37,no.5, pp. 977–997,2004.

5. X. L. Chen, J. Yang, J. Zhang, and A. Waibel, "Automatic detection and recognition of signs from natural scenes," Jan. 2004.

6. K.H.Zhu,F.H.Qi,R.J.Jiang,L.Xu,M.Kimachi,Y.Wu,andT.Aizawa, "UsingAdaboosttodetectandsegmentcharactersfromnaturalsenes,"inProc.1stConf.CarameraBasedDocumentAnalysisand Recognition(CBDAR'05),Seoul,SouthKorea,2005,pp.52-59.

7. ZhangandS.F.Chang,"Learningtodetectscenetextusinga higher orderMRFwithbeliefpropagation,"inProc.IEEEConf.ComputerVisionandPatternRecognition Workshops(CVPRW'04),Washington,DC,2004,pp.101-108.

8. Yi-Feng Pan, XinwenHou, and Cheng-Lin Liu, Senior Member, IEEE,"A Hybrid Approach to Detect and Localize Texts in Natural Scene Images" march 2011

9. Khushbu C. Saner,"RobustApproachtoRecognizeandLocalizeTextfrom NaturalSceneImages,"Oct 2014.

10. S.Audithan, RM. Chandrasekaran,"Document Text Extraction from Document Images Using HaarDiscrete Wavelet Transform,"Vol.36 No.4, pp.502-512, 2009.

