ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 6, June 2015

Detection and Extraction of Text in Images using DWT

Tejashvini R. Hiremath[1], Shivanand M. Patil[2], Dr. V. S. Malemath[3]

Department of CSE, KLE Dr. M.S.S. College of Engineering & Technology, Belagavi, Karnataka, India [1,2,3]

Abstract: Extraction of text from the images is one of the important and challenging tasks. These images can be captured from digital capturing devices, since most of the people use digital cameras and other devices wherein text information is important among all the contents from the images. This has many applications such as reading license plate, helping blind persons, tourists etc. In this paper, an approach to detect and locate the texts in images is presented. Text detection and extraction is hard due to the intricate background, different alignment of text size, orientation in images. Texture-based method and the features of wavelet are used to design the proposed work. In the pre-processing step, the image is converted and divided into 50×50 blocks and then extract texture features. Next the classifier classifies the blocks into text and non-text and finally, merging to have text regions. The proposed method accomplishes a detection rate of 99% on a variety of images.

Keywords: Text Detection, Text Extraction, Texture-based, Wavelet Features

I. Introduction

Humans are surrounded by naturally existing environment which assists him to lead his daily life. With the fast growing technologies, it requires a very little effort to capture this information in the form of digital images using devices such as PDAs, digital cameras, mobile phones etc. The content based texts present in these images can be utilized in various applications such as license plate reading.

Various methodologies have evolved for text extraction namely, region-based, connected-component (CC) based, texture-based etc**.** Region-based method differentiates the text region from the background using color or grayscale properties whereas texture-based method uses textural properties which differentiates text from background. CC-based method uses bottom-up approach which groups smaller components into larger components of the text regions in the image.

However, naturally captured images are likely to have some complexities which pose a challenge in extracting the text forms. These complexities arise due to intricate backgrounds, variations in font size, obscurity in color, positioning of texts, shadow effects and image deformation by low quality capturing devices. These are the hurdles in detection and localization of text in images. In order to overcome these hurdles texture-based method is used in this proposed scheme.

The rest of the paper is organized as follows. Section II briefly reviews the related works. Section III describes briefly about discrete wavelet transform. Section IV gives details of the proposed system. Experimental results are presented in Section V and conclusion in Section VI.

II. Related Work

There are many approaches which have been proposed for the text detection and localization and there are various measures on which these existing work or models can be classified for text detection and text localization.

M. Lyu, J. Q. Song and M. Cai proposed comprehensive method [7] for multiple language video to detect text and to extract this in effective manner, multilingual capacity processing is used and is robust for the complicated background and text alignment. Local thresholding is done for text detection and extraction is by adaptive thresholding, which used region- based method and sobel detector to detect edges because of its isotropic in nature and generate double edges whose density is more than non-text regions and large font size is the demerit and Y. X. Liu, S. Goto and T. Ikenaga proposed contour-based robust algorithm [3] wherein connected component (CC) method and texture-based method is used. This also correctly detects the text in intricate background. The algorithm is based on CC method which extracts all the pixels of text edge and an analysis is done to detect external and internal contours and gradient and geometrical characteristics is examined to construct candidate text regions and classify part non-text regions, this model used the expectation maximization algorithm to binarize each text region and problem here is, this results in maximum false negatives in classification but minimizes classification error and also needs information about segmentation before classifying. X. L. Chen, J. Yang, J. Zhang and A.Waibel proposed automatic detection system [2] which also recognizes the signs which has application in sign translation and includes multiple resolution and scale edge detection. This model effectively handles lighting variations since local intensity normalization method is used and for local features gabor transform and Linear Discriminant Analysis (LDA) method for feature selection but used intensity images for these than using binary information but LDA results in less accuracy rate. Yi-Feng Pan, Xinwen Hou and Cheng-Lin Liu proposed region based method [5] to detect and locate texts in scene images. This method uses the cascade

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 6, June 2015

AdaBoost classifier and other several features for text detection. There is a combination of text lines or words for the text localization in the method known as window grouping method. And further a local binarization method is used to extract the connected components which have text and other non-text components are removed by Markov Random Field (MRF) model, by which text components are localized correctly. The problem with this technique is that cascade classifier used is less accurate and slow in identifying the text components. Jonghyun Park, Toan Nguyen Dinh and Gueesang Lee proposed the method to detect text regions [4] in images of signboards with intricate background. This is for mobile system applications. This method uses both edge profile and fuzzy c-means method for detection of text and the segmentation of regions respectively. It starts to extract all possible object edges with canny edge protector and next the horizontal and vertical direction is done using edge profile analysis on these edge pixels to detect text regions, shop name which exists in signboard. To have candidate text regions, edge profile and geometrical characteristics are tested for each object and classify them from background in an image. Segmentation is done through fuzzy c-means algorithm and it is complicated.

Global features give the prominent results than the local features. From the above description we get to know that region-based method, CC method and others results in poor performance, so texture-based is used in proposed method to have the superior results.

III. DISCRETE WAVELET TRANSFORM

The wavelet transform [18] was developed in 1985 and used in different fields such as image processing, signal processing, video compression and so on. In the frequency domain, an image is decomposed into different components and computed with low-pass and high-pass filters and by one- dimensional discrete wavelet transform (DWT) an image is divided into two parts that is coarse and detailed information. Using two-dimensional DWT an image is divided into four sub-bands namely LL which has coarse information and the other LH, HL and LL have detailed information as shown in the below figure1 and 2 respectively.

For multiple resolution two-dimensional DWT[18] is applied for detection for original image since it has increased processing time and helps in extracting many properties of an image where high-pass filters separates high-frequency part and low-pass separates the low-frequency part. The coarse information is an approximation image which is used for next level of resolution to decompose and other three are horizontal, vertical and diagonal wavelet features and are detailed coefficients of sub-bands which are generated when decomposed. It is performed on rows first then on columns, H and L denote high-pass and low-pass channels respectively and for every decomposition the size of an image is reduced to half. For example, if the sub-band image is produced using low pass filter on columns and high pass filters on rows, its LH sub-band and LL is an approximation image, LH deals with vertical, HL deals with horizontal and HH deals with diagonal features

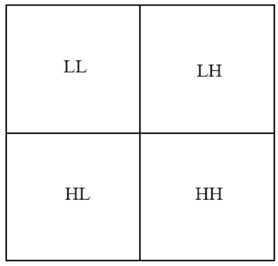

Fig. 1 First-level of decomposition of DWT

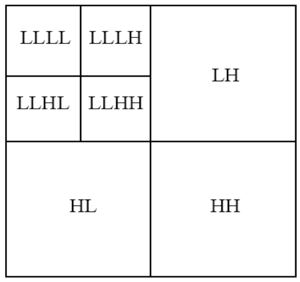

Fig. 2 Second-level of decomposition of DWT

IV. PROPOSED SYSTEM

This section gives an overview of the proposed system. This proposed system uses texture-based method and as we know that there are different textural properties for text and non-text or the background, so that these properties are used to differentiate between them.

Our system consists of four stages as shown in the figure 3 and following are stages.

- Pre-processing
- Feature Extraction
- Classification
- Merging

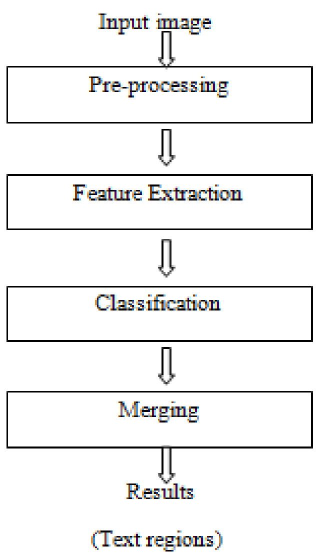

Fig. 3 Flowchart of the proposed system

 DOI 10.17148/IJARCCE.2015.46115

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 6, June 2015

A. Pre-Processing

In this pre-processing stage, the input image is converted into grayscale image or binary image because the aim of this step is to improve the image data that suppresses unwanted distortions or noise which is important for further processing and this is one of the common steps which are carried out to remove noise in the image processing [1].

It also removes the low frequency background noise, various reflections and masks some portions of an image.

The steps which are followed in pre-processing are:

- Input image converted into grayscale image
- Resized to constant size of 300×300 pixels
- Resized image is divided into 50×50 blocks
- Discrete Wavelet Transform (DWT) is applied for each block to get features

B. Feature Extraction

Wavelet energy features are obtained for each 50×50 block in this stage, where image is processed at two resolution levels, from each block we get seven features which are extracted and stored in a variable and it stores the coordinates of the blocks that is minimum and maximum row and column numbers of the blocks. For all the N blocks the values are stored in the variable, which is a feature matrix and where array has all the values of all the blocks and each block consists of values which contains coordinates of row and column and along with this values of sum of all horizontal, vertical and diagonal features of each block, which is computed using summation of all horizontal values to the size of corresponding resolution and horizontal coefficient of wavelet transform is denoted by H, vertical by V and D by diagonal for each block of an image, where H focuses on horizontal edge features, V focuses on vertical edge features and D focuses on diagonal edge features and these have the detailed information which are produced when applied on every resolution and also produce an approximate image which has rough data and used to apply DWT for every resolution and this gives four sub-band images when performed on rows first and then on columns.

These features are used to detect the correct text blocks of images and are fed for classification stage.

C. Classification

This stage classifies every block into text and non-text classes using the discriminant functions and in this case there are only two classes and the classes are denoted by C1 and C2 as text and non-text classes respectively [1]. This classifies the segmented blocks as text or non-text classes using knn classifier based on sample data and training data whose values are stored in separate arrays, which are used in classification.

In order to classify these blocks, k-nearest neighbor (knn) classifier is used which robustly distinguishes these blocks into text and non-text of an image and then coordinates of these respective text blocks of each classified text blocks and stored in an array. Again in this variable or an array we store only the values of the text blocks which are classified out of all the blocks of an image.

These classified text blocks are further processed as inputs to the merging process.

D. Merging

In this stage, the merging process [1] integrates correctly identified text blocks which are stored in an array and the blocks which are connected in rows or columns are checked to have new text regions, and these updated coordinates are stored in variable and has the updated coordinates of text blocks of an image where this new vector has only the coordinates of the classified text blocks and should be equal to the number of text regions of an image. Here in this step input are the values which has only coordinates of text blocks, where the value of the first block is taken to test and assign the new values of coordinates of identified text region to the $block_s$ and then consider the next block from the values in an array and now we check for each block whether it is connected either to a row or a column and if it is connected values of coordinates are updated and merged with the new one, if it's not connected then storing the value of of the region to an array then assign the coordinate values to the new text region of current block these are carried out in iterations till it reaches to the number of text blocks and merged text blocks are displayed as the extracted text region of an image.
Only text region which is extracted and merged is display as the final output.

V. EXPERIMENTAL RESULTS

The proposed system for text region detection and extraction has been tested for 100 natural scene images which have 3600 total blocks and each size is of 50×50 with intricate background and other variations in text font size, color, alignment etc. This method has yielded good results for the images having different text font color, orientation etc. This system accomplishes an overall detection rate of 99% and false reject rate of 0.8% for 100 natural scene images. This system is more favorable as it uses only texture features for text extraction. Several images were tested and among them some of the results are given below.

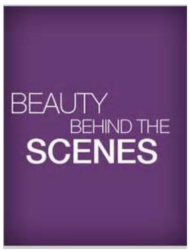

Fig. 4(a) Input image

Fig. 4(b) Extracted Text Region

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 6, June 2015

Fig. 5(a) Input image

Fig. 5(b) Extracted Text Region

Fig. 6(a) Input image

Fig. 6(b) Extracted Text Region

Fig. 7(a) Input image

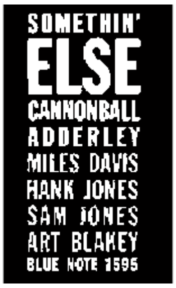

Fig. 7(b) Extracted Text Region

Fig. 8(a) Input image

Fig. 8(b) Extracted Text Region

TABLE I: Overall System Performance

Total no. of blocks tested (# of text blocks)	Correctly detected text blocks	Falsy detected text blocks	Detection rate %
3600 (1562)	1548	14 (0.8%)	99%

VI. CONCLUSION

Texture-based method performs the texture analysis for localization of text from the natural scene images, is presented in this paper. The proposed system is robust and has attained a detection rate of 99% and when compared to connected component method this method is more robust in dealing with the complicated backgrounds and other characteristics of images such as different text font size, color etc. which pose a challenge in extracting text.

However, the proposed method has accomplished better results for natural scene images each of which resized to 300×300 and also detects text regions which are not linearized and finds applications such as helping blind persons, guide tourists', license plate reading etc.

ACKNOWLEDGMENT

Our thanks to respected senior professors, lectures for their support and guidance in completing this project.

REFERENCES

[1] M. M. Kodabagi, S. A. Angadi, Anuradha. R. Pujari, "Text Region Extraction From Low Resolution Display Board Images Using Wavelet Features", International Journal of Information Technology & Management Information System, Vol. 4, Issue 1, January-April (2013), pp. 38-49.

[2] X. L. Chen, J. Yang, J. Zhang, and A.Waibel, "Automatic detection and recognition of signs from natural scenes," IEEE Trans. Image Process., Vol. 13, pp. 87–99, Jan. 2004.

[3] Y. X. Liu, S. Goto, and T. Ikenaga, "A contour-based robust algorithm for text detection in color images," IEICE Trans. Inf. Syst., Volume ol. E89-D, no. 3, pp. 1221–1230, 2006.

[4] Jonghyun Park, Toan Nguyen Dinh, and Gueesang Lee, "Binarization of Text Region based on Fuzzy Clustering and Histogram Distribution in Signboards", World Academy of Science, Engineering and Technology ,Vol:2 2008-07-23.

[5] Yi-Feng Pan, Xinwen Hou, Cheng-Lin Liu," Text Localization in Natural Scene Images based on Conditional Random Field", National Laboratory of Pattern Recognition Institute of Automation, Chinese Academy of Sciences ,10th International Conference on Document Analysis and Recognition,2009.

[6] Xiaopei Liu, Zhaoyang Lu, Jing Li, Wei Jiang," Detection and Segmentation Text from Natural Scene Images Based on Graph Model", State Key Laboratory of Integrated Services Networks 1XiDian University 1 Xi'an 710071 china, WSEAS Transactions on Signal Processing, vol. 10,2014.

[7] M. Lyu, J. Q. Song, and M. Cai, "A comprehensive method for multilingual video text detection, localization, and extraction," IEEE Trans. Circuits Syst. Video Technol., vol. 15, no. 2, pp. 243–255, 2005.

[8] Xu-Cheng Yin, Xuwang Yin, Kaizhu Huang, and Hong-Wei Hao," Robust Text Detection in Natural Scene Images", arXiv:1301.2628v3 [cs.CV] 2 Jun 2013.

[9] Rafael C.Gonzalez, Richard E.Woods, Steven L.Eddins, "Digital Image Processing using MATLAB", Pearson Education.

[10] Hu Shu-jie, Shi Zhen-gang, "A new approach for color text segmentation based on rough set theory", Journal of Convergence Information Technology, Vol. 8, No. 4, Feb. 2013.

[11] Mohanabharathi. R., Surender. K, Selvi. C, "Detecting and localizing color text in natural scene images using Region based and connected component method ", International Journal of Modern Engineering Research, Vol. 3, Issue 1, Jan-Feb. 2013,pp. 331-335.

[12] Miss. Poonam B. Kadam, Mrs. Latika R. Desai, " A hybrid approach to detect and recognize texts in images", International Journal of Research in Computer and Communication Technology, vol. 2, Issue 7, July-2013, pp. 405-410.

[13] Shehzad Muhammad Hanif, Lionel Prevost, " Texture based text detection in natural scene images: A help to blind and visually impaired persons", Conference and Workshop on Assistive Technologies for people with Vision and Hearing Impairments Assistive Technology for all Ages CVHI 2007, M. A. Hersh(ed.).

[14] C. P. Sumathi, T. Santhanam and G. Gayathri Devi, " A survey on various approaches of text extraction in images", International Journal of Computer Science and Engineering Survey, Vol.3, No. 4, August 2012.

[15] Kamarul Hawari Ghazali, Mohd Fais Mansor, Mohd. Marzuki Mustafa and Aini Hussain, " Feature extraction technique using discrete wavelet transform for image classification", The 5th Conference on Research and Development- SCOReD 2007, 11-12 December 2007, Malaysia.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 6, June 2015

[16] Marcin Kociołek1, Andrzej Materka1, Michał Strzelecki1, Piotr Szczypiński1," Discrete Wavelet Transform – Derived Features For Digital Image Texture Analysis", Proc. of Interational Conference on Signals and Electronic Systems, 18-21 September 2001, Lodz, Poland, pp. 163-168.
[17] S. Prabhakar, A.R. Mohanty, A.S Sekhar, "Application of discrete wavelet transform for detection of ball bearing race faults", Tribology International 35, 2002, pp.798-800.
[18] Fang-Hsuan Cheng,Yu-Liang Chen," Real time multiple objects tracking and identification based on discrete wavelet transform", Pattern Recognition 39, 2006, pp.1126 – 1139.
[19] Hasan Ocak, "Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy", Expert Systems with Applications 36, 2009, pp.2027–2036.
[20] Niti Syal, Naresh Kumar Garg," Text Extraction in Images Using DWT, Gradient Method And SVM Classifier", International Journal of Emerging Technology and Advanced Engineering, Vol. 4, Issue 6, June 2014, pp.477-481.
[21] A. Materka, M. Strzelecki,"Texture Analysis Methods – A Review", Technical University of Lodz, Institute of Electronics, COST B11 report, Brussels 1998.

BIOGRAPHIES

Tejashvini R. Hiremath is a M.Tech student in the Department of Computer Science & Engg., KLE Dr. M.S.Sheshgiri College of Engg. & Tech., Belagavi. She received B.E. degree in Computer Science & Engineering from B.L.D.E.A's V.P. Dr. P.G.Halakatti College of Engg. & Tech., Vijaypur in 2013. Her research interests include Image Processing and Wireless Sensor Networks.

Prof. Shivanand M. Patil is a faculty in the Department of Computer Science & Engg., KLE Dr. M.S.Sheshgiri College of Engg. & Tech., Belagavi. He did his Bachelors in Computer Science & Engg. from MMEC, Belagavi and M.Tech in Computer Science & Engg. from Basaveshwar Engineering College, Bagalkot. His research interests include Image Processing and Pattern Recognition. He has number of publications in to his credit in peer reviewed international journals and conferences.

Dr. Virendra. S. Malemath is currently a Head of Computer Science & Engg, KLE DR M S Sheshgiri College of Engg. & Tech., Belagavi. He did his Bachelors in Engg. in Electronics & Communication Engg. from Karnataka University, Dharwad in the year 1993, did his MS in Software Systems from BITS Pilani Rajasthan in 1998 and received his PhD in Computer Science from Gulbarga University, Gulbarga, India in 2009. His research interests are document image processing medical and pattern recognition. He has published more than 60 articles in peer reviewed international journals and conferences.