International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Fast Scene Text Detection with DWT Based Edge Enhancement

Femina Mohammed P[1], Priya K. P.[2]

[1]MTECH Scholar, MEA Engineering College, University of Calicut,
Perinthalmanna, Malapuram, Kerala, India -679325

[2]Assistant Professor, Dept. of ECE,
MEA Engineering College, Perinthalmanna, Malapuram, India

Abstract: ***The detection and extraction of text in natural and camera captured images is always a challenging problem. The text in images contain most of the important information and can be used in a variety of applications like translators for tourists, aids for visually impaired people, in information retrieval systems etc. Implementation of such applications on a general purpose computer can be easier, but every time it is not efficient due to additional constraints on memory and other peripheral devices. The paper discusses an efficient yet simple and fast method to extract text from static images. The speed of Haar discrete wavelet transform is fastest among all wavelets because its coefficients are either 1 or -1. It is one of the reasons that Haar DWT is used to detect edges of candidate text regions. Edges are detected horizontally, vertically and diagonally. A Gaussian filtering is initially applied to obtain a smoothed image. Masking is applied on the image sub-bands obtained by DWT. Then on the IDWT image morphological operations like erosion and dilation are applied to remove the noisy areas along with angle tiltation correction and finally ocr output is obtained.***

Keywords: Discrete Wavelet Transform (DWT), Haar Transform, Edge Enhancement, Erosion, Dilation.

1. Introduction

Due to the increasing use of digital image capturing devices, such as digital cameras, mobile phones and PDAs, great attention have been drawn in content-based image analysis techniques in recent years. Among all the contents in images, text information has inspired great interests, since text in images contain large quantity of useful information which can be used to fully understand image. It also finds wide applications such as license plate reading, sign detection and translation, mobile text recognition, content-based web image search, and so on.

Images captured with a camera have different kinds of degradations. Due to the variety of font size, style, orientation, alignment, as well as the complexity of background, designing a robust general algorithm, which can effectively detect and extract text from images is still a challenging problem.

Hyung Il Koo and Duck Hoon Kim [1] proposed a connected component based text extraction method , which uses two machine learning classifiers. The method consists of three steps- candidate generation, candidate normalization and non text filtering. The extracted CC's have to be partitioned into clusters to form candidate word regions. An Adaboost classifier is trained that determines adjacency relationship and cluster CC's by using their pair wise relationship. The normalized and binarized data is used for text or non-text classification. For the purpose multi layer perceptron is trained. Even though the method yields good performance, it is not fully efficient. The intensity variations of the text affect the detection problem. Also angle aligned text was not able to detect using the method.

Here we are comparing that CC based method with a fast method of text detection which is based on DWT and edge detection. Edge detection and enhancement is done on the Haar wavelet transformed image. Further morphological operations like erosion and dilation are carried out. Finally angle tiltation correction and OCR based noise removal is done.

2. Literature Survey

The text extraction process has gained importance and attraction much earlier. The works discussed here gives a brief description of the previous methods. Paper [2] gives an adaboost algorithm for detecting and reading text in unconstrained city scenes. Adaboost requires specifying a set of features from which a strong classifier is build. A feature set guided by the principle of informative features are selected. The strong classifier is applied to sub regions of image and outputs text candidate regions. The second component an extension and binarization acts on text region candidates, which extends the regions to include text that the strong classifier did not detect and binarizes. The third component is an OCR software program which acts on binarized regions, which determines that the regions are text and reads or rejects them. But even though OCR algorithm is effective, they remain a black box and cannot modify or improve them.

[3] Proposes a CC-based text detection algorithm, which employs Maximally Stable Extremal Regions (MSER) as basic letter candidates. Despite their favourable properties, MSER has been reported to be sensitive to image blur. To allow for detecting small letters in images of limited resolution, the complimentary properties of Canny edges and MSER are combined in the edge-enhanced MSER. Further the paper proposes to generate the stroke width transform

image of these regions using the distance transform to efficiently obtain more reliable results. The geometric as well as stroke width information are then applied to perform filtering and pairing of CCs. Finally, letters are clustered into lines and additional checks are performed to eliminate false positives. The edge-enhanced MSER detected in the query image can be used to extract feature descriptors like for visual search. Further, the system provides a reliable binarization for the detected text, which can be passed to OCR for text recognition.

3. Proposed Method

The proposed method is based on the fact that edges are a reliable feature of text regardless of color/intensity, layout, orientations, etc. Three distinguishing characteristics of text embedded in images are edge strength, density and the orientation, variance, which can be used as main features for detecting text. Block diagram of the proposed method is shown in figure 1. As shown in the figure, the main steps in the method are image smoothing, DWT of the image, masking for edge detection, inverse DWT, connected component extraction, morphological operations like erosion and dilation, angle alignment and OCR output.

3.1 Image Smoothing

Images are basically an array of pixel values and those values are invariably affected by noise. Smoothing reduces the effects of noise by replacing every pixel by a weighted average of its neighbors. Convolving the image with a filter is done in smoothing. Here we are using a Gaussian filter. They are linear low pass filters. Gaussian smoothing is very effective for removing Gaussian noise and their weights are computed according to a Gaussian function. The weights give higher significance to pixels near the edge (reduces edge blurring). The degree of smoothing is controlled by σ (larger σ for more intensive smoothing).

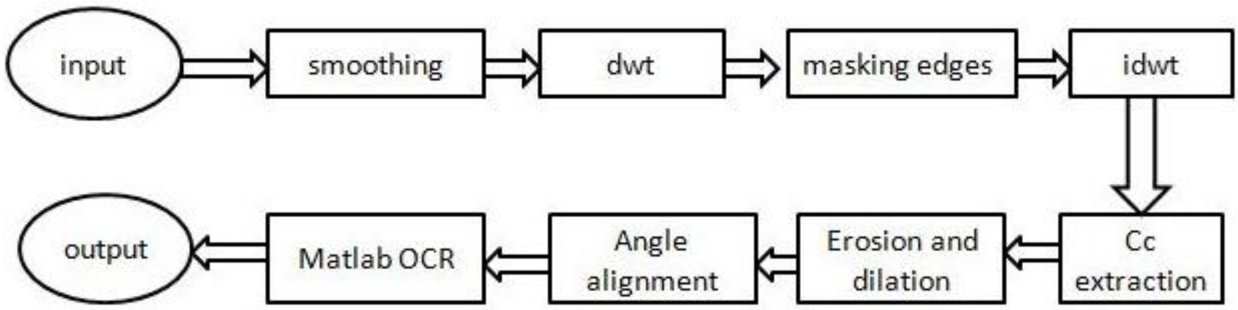

Figure 1: Block Diagram of Proposed System

3.2 Discrete Wavelet Transform (DWT)

Here, we are using Haar discrete wavelet transform which is a powerful tool for modeling the characteristics of textured images. Most textured images are featured by their contained edges. It can decompose signal into different components in the frequency domain. We are using 2-d DWT which decomposes input image into four components or sub-bands, one average component(LL) and three detail components(LH, HL, HH) .The detail component sub-bands are used to detect candidate text edges in the original image. The Haar wavelet, transforms the illumination components to the wavelet domain. This results in the four LL, HL, LH and HH sub image coefficients. The traditional edge detection filters can provide similar results but it cannot detect three kinds of edges at a time. Therefore, processing time of the traditional edge detection filters is slower than 2-d DWT, so we choose Haar DWT.

3.3 Masking the Edges

Now the masking the edges of four sub-bands of the image have to be done. Masking is filtering. We are using four masks for the purpose- a horizontal mask, a vertical mask and two diagonal masks. A convolution operation with a compass operator (as shown in Fig. 2) results in four oriented edge intensity images E(θ), (θ ∈ {0, 45, 90, 135}), which contain all the properties of edges required in our proposed method. Then inverse DWT is applied on the masked bands.

-1	-1	-1
2	2	2
-1	-1	-1

0° kernel

-1	-1	2
-1	2	-1
2	-1	-1

45° kernel

-1	2	-1
-1	2	-1
-1	2	-1

90° kernel

2	-1	-1
-1	2	-1
-1	-1	2

135° kernel

Figure 2: Compass Operator

3.4 CC Extraction

The connected components in the binary image are then obtained. These CCs will contain both text and non text areas. A number of image properties are obtained like area, pixel-list, orientation, major and minor axis length, centroid. In order to remove noisy small CCs, an area range is set. Also to remove the non text portions, the ratio of major and minor axis length is taken. Most of the noise and non text portions will be removed here.

3.5 Dilation and Erosion

These are two important morphological operations applied on binary images. Dilation allows objects to expand, thus potentially filling in small holes and connecting disjoint objects. The dilation process is performed by laying the structuring element **B** on the image **A** and sliding it across the image in a manner similar to convolution. If the origin of the structuring element coincides with a 'white' pixel in the image, there is no change; move to the next pixel. If the origin of the structuring element coincides with a 'black' in the image, make black all pixels from the image covered by the structuring element. Erosion shrinks objects by etching away (eroding) their boundaries.

The erosion process is similar to dilation, but we turn pixels to 'white', not 'black'. If the origin of the structuring element coincides with a 'white' pixel in the image, there is no change; move to the next pixel. If the origin of the structuring element coincides with a 'black' pixel in the image, and at least one of the 'black' pixels in the structuring element falls over a white pixel in the image, then change the 'black' pixel in the image (corresponding to the position on which the center of the structuring element falls) from' black' to a 'white'.

3.6 Angle Alignment and OCR output

For finding the angle tiltation of the text, properties like centroid and orientation are made use of. An important feature of this method is the angle tiltation correction, which most of the text detection algorithms will not make use of. With respect to the centroid, the angle tilted can be easily found out and is corrected.

The goal of Optical Character Recognition (OCR) is to classify optical patterns (often contained in a digital image) corresponding to alphanumeric or other characters. The processed data is given to matlab OCR to display the detected text portions from the image.

4. Results and Discussion

The results of the two methods are compared here- the CC based and the DWT based text extraction method. Various parameters like precision, recall, f measure and accuracy of the two methods are estimated for better comparison.

Figure 3: Original Image

Figure 3 shows a camera image containing both text and non-text regions. The text portions extracted by both the algorithms are shown in the following figures.

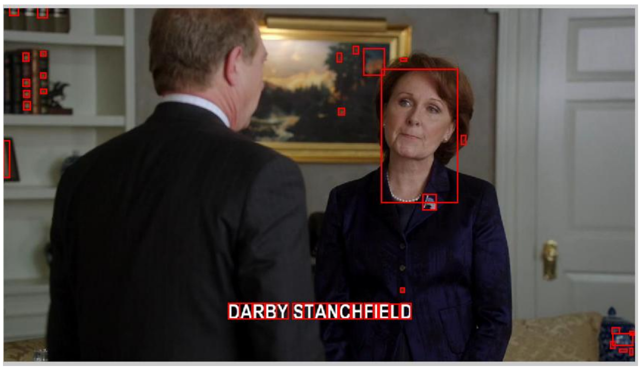

Figure 4: Output of CC Based Method

Figure 5: Output of Proposed Method

By comparing figure 4 and 5 it can be seen that, even though the CC based method extracts the text portions in the image, along with that, it extracts many non-text portions also(indicated by the red boxes). Also all of the text portions are not detected, there are some missing letters.

On the other hand, the proposed DWT based method efficiently extracts all of the text portions (shown in the yellow box) with no non-text component in it. The matlab OCR result shown in figure 6 also shows the text output obtained by this method.

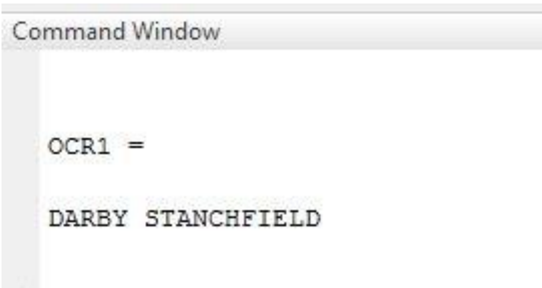

Figure 6: OCR output

5. Conclusion

Text detection is very challenging problem now-a-days and efficiently detecting text from images is a complex process. The paper proposes a new and fast scene text detection algorithm which makes use of dwt along with edge detection and edge enhancement. Most of the shape information of an image is enclosed in edges. So detecting the edges and enhancing areas containing edges will increase the sharpness of an image and can efficiently detect the text portions. The paper also compares the results of this proposed method with an existing CC based text extraction method. It was shown that the accuracy and efficiency of proposed method is higher than the existing one.

References

[1] Hyung Il Koo, Member, IEEE, and Duck Hoon Kim, Member, Scene Text Detection via Connected Component Clustering and Non-text Filtering IEEE; IEEE transactions on image processing, vol. 22, no. 6, june 2013
[2] X. Chen and A. Yuille, “Detecting And Reading Text In Natural Scenes,”In Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2004,Pp. 366–373
[3] H. Chen, S. Tsai, G. Schroth, D. Chen, R. Grzeszczuk, And B. Girod, “Robust Text Detection In Natural Images With Edge-Enhanced Maximally Stable Extremal Regions,” In Proc. IEEE Int. Conf. Image Process., Sep. 2011, Pp. 2609–2612.
[4] C. Yao, X. Bai, W. Liu, Y. Ma, And Z. Tu, “Detecting Texts Of Arbitrary Orientations In Natural Images,” In proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2012, Pp. 1083–1090
[5] B. Epshtein, E. Ofek, And Y. Wexler, “Detecting Text In Natural Scenes With Stroke Width Transform,” Inproc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2010, Pp. 2963–2970
[6] A. Shahab, F. Shafait, and A. Dengel, “Icdar2011 Robust Reading Competition Challenge 2: Reading Text in Scene Images,” In proc. Int. Conf. Document Anal. Recognit., 2011, Pp. 1491–1496