Text Extraction and Recognition from an Image Using Image Processing In Matlab

Adesh Kumar
Electrical, Electronics and Instrumentation Department
College of Engineering Studies,
University of Petroleum and Energy Studies, Dehradun, 248001, India

Pankil Ahuja
Student, B.tech. (Electronics)-4th year
College of Engineering Studies,
University of Petroleum and Energy Studies, Dehradun, 248001, India
Email: pankilahuja@gmail.com
www.upes.ac.in

Rohit Seth
Student, B.tech. (Electronics)-4th year
College of Engineering Studies,
University of Petroleum and Energy Studies, Dehradun,248001, India

Abstract

This project extracts the text from any image, including the famous number plate extraction and recognizes the text written. Text Extraction and recognition in general have quite a lot of relevant application for automatic indexing or information retrieval such document indexing, content-based image retrieval, and the famous car plate recognition which further opens up the possibility for more improved and advanced systems. “Haar transform” is strictly used in this approach of extracting text which can be further use into many applications.

Keywords: Haar DWT, Sobel edge detector, AND operator, thresholding

1. INTRODUCTION

Extracting texts explores the possibility for more advanced systems, this project provides the flexibility to extract texts and show them line-by-line (in case more that a line extracted). It includes the "Haar transform" followed by converting the image into black and white format with some threshold value because the black and white comprises only of [0, 1].

A visual image is rich in information. The main purpose of image enhancement is to bring out detail that is hidden in an image or to increase contrast in a low contrast image. Image enhancement is among the simplest and most appealing areas of digital image processing. Basically, the idea behind enhancement techniques is to bring out detail that is obscured or simply to highlight certain features of interest in an image. The detection and extraction of text regions in an image is a well known problem in the computer vision research area. Compressing an image is significantly different than compressing raw binary data. General purpose compression programs can be used to compress images, but the result is less than optimal. This is because images have certain statistical properties which can be exploited by encoders specifically designed for them. Also, some of the finer details in the image can be sacrificed for the sake of saving a little more bandwidth or storage space. This also means that lossy compression techniques can be used in this area.

The discrete wavelet is essentially sub band-coding system and sub band coders have been quite successful in speech and image compression. In this paper, we have implemented HAAR Wavelet Transform which shows that the Haar transform can be used for image compression. There is an efficient yet simple method to extract text regions from static images or video sequences. The operation speed of Haar discrete wavelet transform (DWT) operates the fastest among all wavelets because its coefficients are either 1 or -1. This is one of the reasons we employ Haar DWT to detect adges of candidate text regions. The resulted detail component sub-bands contain both text edges and non-text edges. Therefore, we can apply thresholding preliminary remove the non-text edges. Text regions are composed of vertical edges, horizontal edges and diagonal edges.

Morphological dilation operators are applied to connect isolated text edges of each detail component sub-band in a transformed binary image. According to the experiment results, real text regions are the overlapped portion of three kinds of dilated edges. Hence, we can apply the logical OR operator to the three kinds of dilated edges and obtain the final text regions correctly.

2. Haar discrete wavelet transform

The discrete wavelet transform is a very useful tool for signal analysis and image processing, especially in multi-resolution representation. It can decompose signal into different components in the frequency domain. One-dimensional discrete wavelet transform (1-D DWT) decomposes an input sequence into two components (the average component and the detail component) by calculations with a low-pass filter and a high-pass filter[2]. Two-dimensional discrete wavelet transform (2-D DWT) decomposes an input image into four sub-bands, one average component (LL) and three detail components (LH, HL, HH) as shown in Figure 1. In image processing, the multi-resolution of 2-D DWT has been employed to detect edges of an original image. The traditional edge detection filters can provide the similar result as well. However, 2-D DWT can detect three kinds of edges at a time while traditional edge detection filters cannot.

As shown in Figure 2, the traditional edge detection filters detect three kinds of edges by using four kinds of mask operators. Therefore, processing times of the traditional edge detection filters is slower than 2-D DWT.

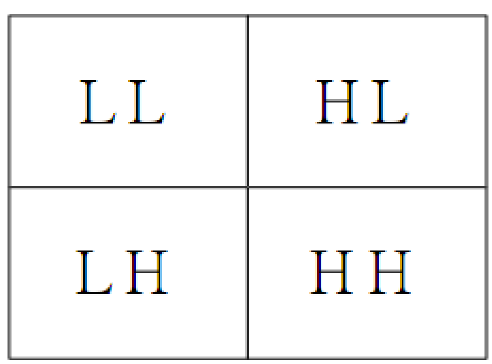

Figure 1 shows a gray level image

As we can see, three kinds of edges present in the detail component sub-bands but look unobvious (very small coefficients). If we replace the taps DWT filters with Haar DWT, the detected edges become more obvious and the processing time decreases. The operation for Haar DWT is simpler than that of any other wavelets. It has been applied to image processing especially in multi-resolution representation[3]. Harr DWT has the following important features[4].

1. Haar wavelets are real, orthogonal, and symmetric.
2. Its boundary conditions are the simplest among all wavelet-based methods.
3. The minimum support property allows arbitrary spatial grid intervals.
4. It can be used to analyze texture and detect edges of characters.
5. The high-pass filter and the low-pass filter coefficient is simple (either 1 or –1).

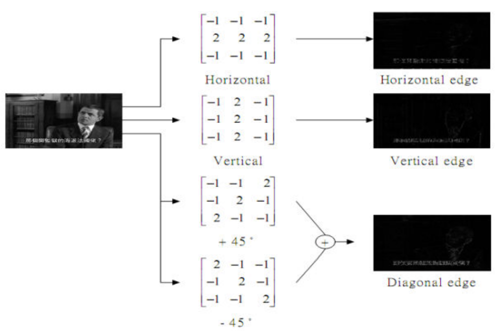

Figure 2. Traditional edge detection using mask operation.

Figure 3 (a) shows the example of a 4×4 gray-level image. The wavelet coefficients can be obtained in gray-level image using addition and subtraction. 2-D DWT is achieved by two ordered 1-D DWT operations (row and column). First of all, we perform the row operation to obtain the result shown in Figure 3 (b). Then it is transformed by the column operation and the final resulted 2-D Haar DWT is shown in Figure 3 (c). 2-D Haar DWT decomposes a gray-level image into one average component sub-band and three detail component sub-bands.

$$\begin{bmatrix} A & B & C & D \\ E & F & G & H \\ I & J & K & L \\ M & N & O & P \end{bmatrix} \qquad \begin{bmatrix} (A+B) & (C+D) & (A-B) & (C-D) \\ (E+F) & (G+H) & (E-F) & (G-H) \\ (I+J) & (K+L) & (I-J) & (K-L) \\ (M+N) & (O+P) & (M-N) & (O-P) \end{bmatrix}$$

(a) (b)

$$\begin{bmatrix} (A+B)+(E+F) & (C+D)+(G+H) & (A-B)+(E-F) & (C-D)+(G-H) \\ (I+J)+(M+N) & (K+L)+(O+P) & (I-J)+(M-N) & (K-L)+(O-P) \\ (A+B)-(E+F) & (C+D)-(G+H) & (A-B)-(E-F) & (C-D)-(G-H) \\ (I+J)-(M+N) & (K+L)-(O+P) & (I-J)-(M-N) & (K-L)-(O-P) \end{bmatrix}$$

(c)

Figure 3 (a) the original image (b) the row operation on 2D Haar DWT (c) the column

2.1 Operation of 2-D Haar DWT

In those three detail components of a Haar DWT image, we can obtain various features about the original image as follows:
1. Average components are detected by the LL sub-band;
2. Vertical edges are detected by the HL sub-band;
3. Horizontal edges are detected by the LH sub-band;
4. Diagonal edges are detected by the HH sub-band.

Figure 4. 2-D Haar discrete wavelet transform image

We can detect candidate text edges in the original image from those three detail component sub bands (HL, LH and HH) in Figure 4.

Chen and Liao[5] presented the segment-matrix algorithm for Haar DWT to decrease the processing time of DWT operations. The method produces the same results as traditional Haar DWT with a much faster speed. Hence, we apply the segment-matrix algorithm to decompose an original gray-level image into four sub-bands. After the Haar DWT, the detected edges include mostly text edges and some non-text edges are presented in the detail component subbands. In next subsection, we employ dynamic thresholding to preliminarily remove those non-text edges in the detail component sub-bands.

2.2 Thresholding

Thresholding is a simple technique for image segmentation. It distinguishes the image regions as objects or the background. Although the detected edges are consist of text edges and non-text edges in every detail component sub-band, we can distinguish them due to the fact that the intensity of the text edges is higher than that of the non-text edges. Thus, we can select an appropriate threshold value and preliminarily remove the non-text edges in the detail component sub-bands. In this subsection, we employ dynamic thresholding[6] to calculate the target threshold value T. The target threshold value is obtained by performing an equation on each pixel with its neighboring pixels. We employ two mask operators to obtain such an equation and then calculate the threshold value for each pixel in the detail sub-bands. Basically, the dynamic thresholding method obtains different target threshold values for different images. Each detail component sub-band is then compared with T to obtain a binary image.The threshold T is determined by where

$$T = \frac{\sum (es(i,j) \times s(i,j))}{\sum s(i,j)} \qquad (2)$$

$$s(i,j) = Max(|g1 ** es(i,j)|, |g2 ** es(i,j)|) \qquad (3)$$

$$g1 = [-1 \quad 0 \quad 1], g2 = [-1 \quad 0 \quad 1]^t \qquad (4)$$

In Eq. (3), denote two-dimensional liner convolution.

Figure 5 shows the example of a 5×5 detail component sub-band. We calculate S(P8) as an example to demonstrate the definition of Eqs. (3) and (4).

$$S(P8) = max(|P9 - P7|, |P13 - P3|) \qquad (5)$$

Applying similar operations to each pixel, we obtain all the S(i, j) for each detail component sub band. After that, we can apply Eq. (2) to compute T and the binary edge image (e) is then given by

$$e(i,j) = \begin{cases} 255, & if\ es(i,j) > T \\ 0, & otherwise \end{cases} \qquad (6)$$

$$\begin{bmatrix} P1 & P2 & P3 & P4 & P5 \\ P6 & P7 & P8 & P9 & P10 \\ P11 & P12 & P13 & P14 & P15 \\ P16 & P17 & P18 & P19 & P20 \\ P21 & P22 & P23 & P24 & P25 \end{bmatrix}$$

Figure 5 A detail component 5×5 sub-band (s)

2.3 Text region extraction

In this subsection, we use morphological operators and the logical AND operator to further removes the non-text regions. In text regions, vertical edges, horizontal edges and diagonal edges are mingled together while they are distributed separately in non-text regions. Since text regions are composed of vertical edges, horizontal edges and diagonal edges, we can determine the text regions to be the regions where those three kinds of edges are intermixed. Text edges are generally short and connected with each other in different orientation. In Figure 6, we use different morphological dilation operators to connect isolated candidate text edges in each detail component sub-band of the binary image. In this research, 3×5 for horizontal operators, 3×3 for diagonal operators and 7×3 for vertical operators as in shown Figure 7 are applied. The dilation operators[9] for the three detail sub-bands are designed differently so as to fit the text characteristics. The logical AND is then carried on three kinds (vertical, horizontal and diagonal) of edges after morphological dilation. This process is indicated in Figure 8. Since three kinds of edge regions are intermixed in the text regions, overlapping appears a lot after the morphological dilation due to the expansion of each single edge. On the contrary, only one kind of edge region or two kinds of edge regions exist separately in the non-text regions and hence there is no over-lapping even after the dilation. Therefore, the AND operator helps us to obtain the candidate text regions. Some-times the text candidate regions may contain some non-text component regions which are too large or too small. By limiting the block size, we obtain the final text regions. Each text region has a moderate size w × h (pixels) in a candidate text region image

Figure 6. The dilated image of three binary regions

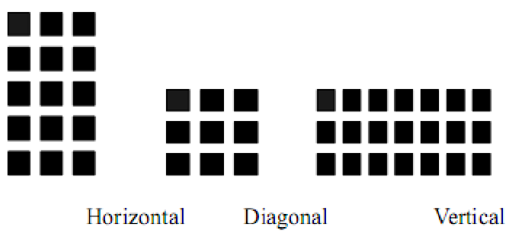

Figure 7. Horizontal, Diagonal and Vertical edges dilation operators

The minimum text block size is determined as follows

width>100pixels,height>35(pixel) (7)

Removing the candidate text regions smaller than this limit, the final text region is shown in Figure 9 (b).

And therefore, text written in any language can be easily extracted and can be use for further application purpose.

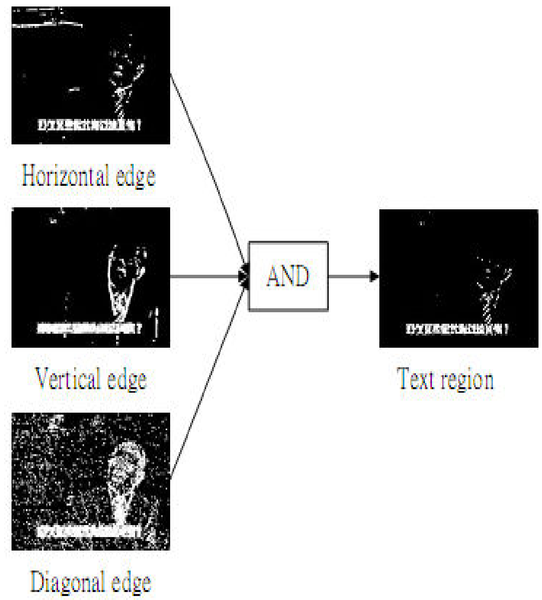

Figure 8. Text extraction by using the logical AND operator

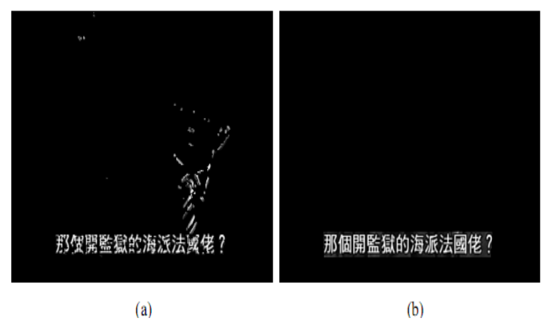

Figure 9 (a) The candidate text region (b) the extracted real text region

The dense edges are the distinct characteristics of the text blocks which are used to detect the possible text regions. The candidate text regions are found by finding the edges in the mentioned sub-bands and fusing the edges contained in each sub-band. The Sobel edge detector[10] is efficient to extract the strong edges. The Sobel edge detector is applied on each sub-band to get the candidate text edges. In the next step, using a weighted 'OR' operator, these candidate text edges are used to form the edge map. A threshold is applied on the edge map to obtain the binary edge map. Then, a morphological dilation operation is performed on the processed edge map. This operation results in filling the gaps inside the obtained characters' regions.

2.4 Removing Non-Text Regions

To further enhance the results, the non-text regions are removed. The horizontal rectangular areas with high density indicate text strings. Projection is a more efficient way to find such high density areas. The idea of coarse-to-fine detection is to locate the text region progressively by two phase projection. There are lots of detected edge dense blocks that include multi-line texts. The projection profile is used to separate these blocks into single line text. A horizontal/vertical projection profile is defined as the vector of the sums of the pixel intensities over each column/row[9]. The horizontal and vertical projection of the processed edge map is found. The average of the minimum and maximum value of the vertical projection is taken as the threshold. Then the rows whose sums of pixel intensities are above the threshold are taken. Next, the horizontal projection of only those rows is found. The minimum and maximum of the horizontal projection is taken and the average of them is taken as the threshold. Only the columns whose sums of pixel intensities are above the threshold are taken. This results in the localization of the text regions in the image. The corresponding regions in the original grayscale image are taken. Finally, a threshold is applied on these regions, which results in the segmentation of the real text regions from the image.

3. Conclusion

A method of text extraction from images is proposed using the Haar Discrete Wavelet Transform, the Sobel edge detector, the weighted OR operator, thresholding and the morphological dilation operator. These mathematical tools are integrated to detect the text regions from the complicated images. The proposed method is robust against language and font size of the texts. The proposed method is also used to decompose the blocks including multi-line texts into single line text. According to the experimental results, the proposed method is proved to be efficient for extracting the text regions from the images.

4. References

[1] SEGMENTATION OF TEXT FROM COMPOUND IMAGES by Dr.N.Krishnan , C. Nelson Kennedy Babu , S.Ravi and Josphine Thavamani presented in *International Conference on Computational Intelligence and Multimedia Applications 2007*.

[2] Qixiang Ye, Qingming Huang, Wen Gao and Debin Zhao, Fast and *Robust text detection in images and video frames, Image and Vision Computing 23, 2005.*

[3] Rainer Lienhart and Axel Wernicke, Localizing and Segmenting Text in Images and Videos, *IEEE Transactions on Circuits and Systems for Video Technology, Vol.12, No.4, April 2002.*

[4] Fujii, Masafumi., Wolfgang, J. R., and Hoefer. 2001. Filed-Singularity correction in 2-D time-domain Haar-wavelet modeling of waveguide components. *IEEE Transactions on Microwave Theory and Techniques, 49, 4*.

[5] Chen, P. Y. and Liao, E. C. 2002. A new algorithm for Haar discrete wavelet transform. *IEEE International Symposium on Intelligent Signal Processing and Communication Systems, 21, 24: 453-457.*

[6] Park, C. J., Moon, K. A., Oh, Weon- Geun, and Choi, H. M. 2000. An efficient of character string positions using morphological operator. *IEEE International Conference on Systems, Man, and Cybernetics, 3, 8-11: 1616-1620.*

[7] Xiaoqing Liu and Jagath Samarabandu, Multiscale edge-based *Text extraction from Complex images, IEEE, 2006.*

[8] Matlab version 7.11.0.584 (R2010b) license no. 161051

[9] R.C. Gonzales and R.E. Woods, *Digital Image Processing*, Addison-Wesley, Reading, 1992.

[10] J.D. Foley, A. van Dam, S.K. Feiner and J.F. Hughes, *Computer Graphics, Principles and Practice*, Addison-Wesley, Reading, 1990.