

A Wavelet based Method for Text Segmentation in Color Images

Priya.M

Pg Scholar, Department of EEE,
SNS College of Engineering,
Coimbatore, India

Ck.Gobu

Assistant Professor, Department of EEE,
SNS College of Engineering,
Coimbatore, India

ABSTRACT

This is a digital era, in which imaging has become simple and easy with so many handheld compact devices. Extraction of the text from captured images also becomes necessary for efficient indexing and retrieval purpose. But real time images have issues such as sensor noise, blur, viewing angle, different font size, low contrast etc. Though so many efficient methods exist, they faultier when it comes to complex background images. A wavelet based method is used here. This method preprocesses the image before it is given to an OCR filter. A color quantizer is used to minimize the number of distinct colors in the input image. A discrete wavelet transform is performed on the color quantized image which classifies the image into text and non-text pixels based on their color and the standard deviation of the wavelet. This step is followed by fuzzy C means clustering which partitions the image into the background and text regions. This preprocessed image is then passed through an OCR filter to check the quality of text being segmented.

Keywords- Discrete Wavelet Transform, Fuzzy C-Means Clustering, OCR, Segmentation.

1. INTRODUCTION

Optical Character Recognition (OCR) is a process that converts images of scanned texts or pictures with characters into machine editable text. A pdf file or a scanned image of a magazine can be converted to editable text. Optical character recognition using optical techniques such as mirrors and lenses and digital character recognition using scanners and computer algorithms were originally considered separate fields. OCR has been broadened to include digital image processing. Early systems required training to read a specific font. Intelligent systems with a high degree of accuracy for most fonts are very common nowadays. Some systems are even capable of reproducing formatted output that closely approximates the scanned page including images, columns and other non-textual components. However the approach is sensitive to the size of the font and the font type. For handwritten input, the task becomes even more formidable. Soft computing has been adopted in the process of character recognition for its ability to create input output mapping with a good approximation. The alternative for input output mapping may be the use of a lookup table that is totally rigid with no room for input variations.

Text appearing in images can be classified into two groups: scene text and artificial text. Scene text is part of the image and does not represent any information about the image content, whereas artificial text is laid over the image in a later stage. Artificial text is usually a good key to index images or videos. To obtain such useful indexing data, an OCR system must be used. If the original images are fed directly into an OCR system, the OCR results are often not sufficient due to the complexity of

the image background. Text segmentation is aimed at simplifying or removing the background and thus increasing the text quality in order to achieve good results with OCR systems.

Image pre-processing is the term for operations on images at the lowest level of abstraction. These operations do not increase the image information content but they decrease it if entropy is an information measure. The aim of pre-processing is to improve the image data by suppressing undesired distortions or enhances some relevant image features for further processing and analysis task. Image pre-processing uses the redundancy in images. Neighboring pixels corresponding to one real object have the same or similar brightness value. If a distorted pixel can be picked out from the image, it can be restated as an average value of neighboring pixels. Image pre-processing methods can be classified into categories according to the size of the pixel neighborhood that is used for the calculation of a new pixel brightness.

This paper involves text detection, localization and extraction of text from images with complex background. The OCR filtering stage is not included in this system. The color of the text and background is determined by color quantization. This decreases the distinct number of colors present in the image which will relatively simplify the input image. This simplified image is further decomposed by performing a wavelet transform on it. The color of the pixel and the standard deviation of the wavelet will distinguish the pixels into text and non-text pixels. The text pixels are further classified by Fuzzy C-means clustering (FCM) algorithm. FCM clusters the image into foreground and background clusters. This clustering information is used to produce a binary image with black text on white background. ICDAR data sets have been used to evaluate this proposed method.

2. PREVIOUS WORK

S.Antani [1] has proposed a shot segmentation approach followed by recognition of text for video indexing. The image is binarized, two segmented text regions are generated. The connected components are determined and those which does not satisfy the required aspect ratio are eliminated. The segmented text regions are given a score based on their text-like characteristics. The image with the maximum score goes into the OCR filtering stage. Text detection methods are classified into two, texture based and region based methods. [3,9,19] are text based methods where images are scanned at a number of scales. The neighborhood pixels are classified according to high density edges, DCT Coefficients, distribution of wavelet, high variance of intensity etc. Lack of precision, computational complexity, problems with information integration are the limitations of this method. Region based methods are discussed in [6,10]. Pixels exhibiting certain properties such as approximately constant color are grouped together. The connected components are filtered geometrically and using texture properties to remove unwanted components.

This method simultaneously detects text at any scale and is not limited to horizontal texts.

An unsupervised text segmentation method by Julinda Gllavata [4] uses a clustering algorithm to cluster the image as foreground and background, resulting in a binarized image with the text in black color and the background in white color. This method claims to have achieved better results in the case of images with complex backgrounds. Bunke [16] clustered text pixels from images using a common image segmentation or color clustering algorithms. Heuristic analysis of local connectivity of handwritten text by Kamel [5] and Milewski and Govindaraju [12] addresses the problem of varying illumination with an image by searching for stroke locations and targeting only nonstroke areas. While Kamel [5] removes the noise in nonstroke areas by interpolation and thresholding Milewski and Govindaraju [12] restores the strokes of handwritten forms using a simple interpolation of neighboring pixels. The spatial constraints applied to the images are determined by a heuristic. Niblack [13] and Sauvola [15] used histogram analysis to determine the threshold locally. Lee and Kankanhalli [8] use a combined detection and segmentation stage. Detection is performed by searching for strokes with the same gray level. Each potential character is thresholded using the gray level of its boundary. Le Bourgeois [7] assumes that the dominant portion of the image histogram is the background. Binarization is performed using a maximum entropy thresholding scheme with an additional stage to split characters inadvertently connected by the threading. Messelodi and Modena [9] use a simple global thresholding scheme at the tails of the histogram to extract text from book covers.

Zhang et al [20] proposed a two step iterative conditional random field algorithm with a BP interference. Two types of neighborhood relationships are used. This method effectively extracted multiple text lines . Y.F Pan [14] presented a CRF model for component analysis using text confidence as a unary feature. D.Q. Zhang [19] proposed a parts based approach for 3D scene text detection using a three order MRF model. Wu.et.al [18] use a low pass Gaussian filter to smooth the image and compute an intensity histogram. The histogram is also smoothed and the first peak from the left of the histogram is used as a threshold for the binarization process. Wolf.et.al [17] proposed a text localization, enhancement and binarization method for multimedia documents. A local thresholding is calculated for each block separately.

3. TEXT SEGMENTATION

The method proposed in this paper is designed to segment horizontally aligned scene text and artificial text of any font, size and color. The segmentation module operates on the bounding boxes in the image. The output of segmentation is a binary image of black text pixels on a white background.

The segmentation approach is divided into the following steps:

1. Image Enhancement
2. Color Quantization
3. Discrete Wavelet Transforms
4. Fuzzy C means Clustering
5. Morphological Filtering

3.1 Image Enhancement

Image enhancement alters the visual impact on the interpreter . It improves the information content of the image. Most commercial OCRs performs best only if the images have a resolution of 300 dpi. Not all images have a resolution of 300 dpi. Hence in order to improve the resolution of such images a cubic interpolation is performed on the image. The image now qualifies to be acted upon by the subsequent steps involved in segmentation.

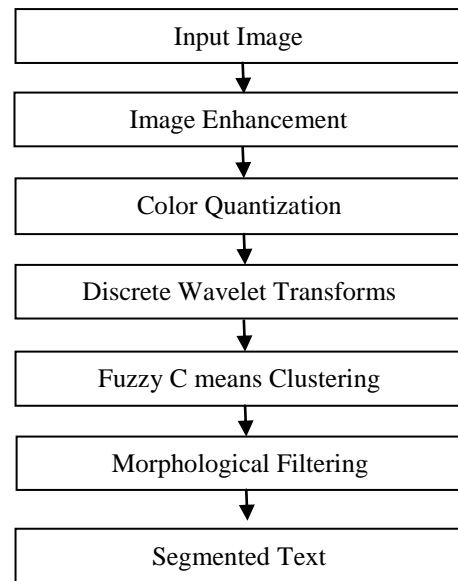


Figure 1. Flowchart of the proposed method

3.2 Color Quantization

Color quantization reduces the number of colors present in the input image. The resulting image has only the most dominating colors present in it. But to us it appears to be the same input image. Color quantization is performed by a color quantization method [18]. Two color histograms are constructed and the difference histogram is calculated. The maximum of this difference histogram gives the text color and the minimum gives the background color.

3.3 Discrete Wavelet Transform

Wavelet analysis is similar to the Fourier analysis. Wavelets are classified as linear transform that is capable of displaying the transformed output at multiple resolutions depending on the point of time/space and at the desired frequency. The wavelet transform was developed to overcome the shortcoming of the Short Time Fourier Transforms. To process images at multiple resolutions Discrete Wavelet Transform (DWT) is the obvious choice. DWT, which is based on sub-band coding yields a faster wavelet transform. It is possible to reduce the computation time and resources required. DWT can be implemented using filter banks consisting of high pass and low pass filters. The wavelet transform to an image consist of a filtering process in the horizontal direction and a subsequent filtering process in the vertical direction.

There are numerous basic functions called mother wavelets that can be used based on the application. Since the mother wavelet produces all wavelet functions used in the transformation through translation and scaling , it determines the characteristics of the resultant wavelet . A Haar wavelet has been used here.

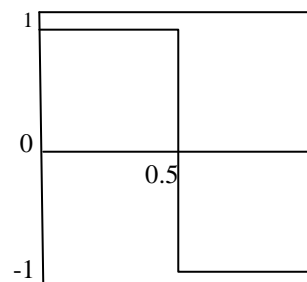


Figure 2. Haar Wavelet

The mother wavelet function $\psi(t)$ of a Haar Wavelet is described as:

$$\psi(t) = \begin{cases} 1 & 0 < t \leq 1/2 \\ -1 & 1/2 \leq t < 1 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Its scaling function $\varphi(t)$ is described as:

$$\varphi(t) = \begin{cases} 1 & 0 \leq t < 1 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

The standard deviation of the wavelet coefficients in the sliding window is expected to be low within a character's texture and high at its boundaries. The standard deviation of wavelet coefficients in a sliding window at position (x,y) is defined as follows:

$$stdev_{window}(x, y) = \sqrt{\frac{1}{N^2} (\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (I(x+i, y+j) - mean_{window})^2)} \quad (3)$$

where $I(x+i, y+j)$ is the wavelet coefficient at pixel position $(x+i, y+j)$

$$mean_{window}(x, y) = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} I(x+i, y+j) \quad (4)$$

3.4 Fuzzy C-means Clustering

Fuzzy C-means Clustering is one of the most efficient algorithms for image segmentation. It allows a pixel to exist in more than one cluster and the degree of membership of that pixel can be identified. This helps to segment the objects in an image effectively and accurately. The identification of the pixel membership in each cluster helps to segment the overlapped objects accurately. The number of clusters is decided by the number of objects present in the image. The centroid is chosen by computing the mean of the data point or randomly. The exact centroid is obtained after several iterations until the coefficients have converged. The centroid is calculated as follows :

$$c_k = \frac{\sum_x w_k(x)x}{\sum_x w_k(x)} \quad (5)$$

where $w_k(x)$ is the degree of belonging to the k^{th} cluster. Here there are two clusters, one is the foreground and the other is the background. The background is eliminated and only the foreground that contains the text region is segmented.

3.5 Morphological Filtering

The basic morphological filtering operations such as erosion and dilation are performed to remove unwanted small objects from the image.

4. SIMULATION AND RESULTS

The data for the tests are obtained from ICDAR 2003 robust reading competition. The data sets had varying challenges such as low resolution, uneven lighting, complex backgrounds, sensor noise etc. Further noise such as Gaussian, Salt and Pepper, Speckle, Poisson were added to the images to find out the precision rate under these noisy conditions. The simulations were repeated with different types of blur added to the test data. The

simulation results are given below and the precision rate is compared with other existing methods.



(a)

THE
RAB BUTLER
BUILDING

(b)



(a)

EMERGENCY
STOP

(b)



(a)

PEPSI

(b)

Figure 3. Text Segmentation Results: (a) Input Image (b) Results of this segmentation method

The Recall rate and the Precision rates are calculated as follows:

$$Precision\ rate = \frac{N_r}{N} \quad (6)$$

$$Recall\ rate = \frac{N_e}{N} \quad (7)$$

where N is the total number of characters, N_r is the number of correctly recognized characters and N_e is the total number of extracting characters.

Table 1. Performance comparison with other methods

Method	Precision Rate (%)	Recall Rate (%)
Proposed Method	82.7	87.6
CRF	56.7	56.9
Y.F Pan	71	67

5. CONCLUSION

In this work we have applied a discrete wavelet transform to decompose the image and then classified the image into the background and foreground using fuzzy c means clustering algorithm. The segmented foreground consists of the text. The unwanted regions are removed by morphological filtering. The proposed method is reliable and robust to the complex background and geometric variations. This method is applicable to both artificial and scene text images. This work can be further extended for video images.

6. REFERENCES

- [1] S. Antani, D. Crandall, R. Kasturi 2000. Robust extraction of text in video. Proceedings of the International Conference on Pattern Recognition (ICPR' 00) IEEE.
- [2] Bezdek, James C. 1981. Pattern Recognition with Fuzzy Objective Algorithms ISBN 0-306-40671-3.
- [3] Chen X, Yuille A, 2004. Detecting and reading text in natural scenes. Proceedings of 2004 IEEE Computer Society Conference On Computer Vision And Pattern Recognition.
- [4] A.Jain, B. Yu, 1998. Automatic Text Location in Images and Video Frames. Pattern Recognition 31(12):2055-2076
- [5] M. Kamel, A. Zhoa, 1993. Extraction of Binary Characters/ Graphic images from gray scale document images. CVGIP: Graphics model image processing , Vol.55, no.3.
- [6] H.K Kim,1996. Efficient automatic text location method and content-based indexing and structuring of the video database. J vi's Commun Image Represent 7 (4): 336-344
- [7] F. LeBourgeois,1997. Robust multifront OCR system from Gray Level Images. International Conference on Document Analysis and Recognition, volume 1.
- [8] C.M. Lee and Kankanhalli 1995. Automatic extraction of characters in Complex Scene Images. International Journal of Pattern Recognition and Artificial Intelligence, 9(1):67-82.
- [9] R. Lienhart, A. Wernicke, 2002. Localizing and Segmenting Text in Images and Videos. IEEE Transactions On Circuits And Systems For Video Technology, Vol.12,NO.4.
- [10] Y.Liu, S.Goto, T.Ikenaga, 2006. A Contour-Based Robust Algorithm for Text Detection in Color Images.IEICE TRANS. INF. & SYST., VOL.E89-D, NO.3.
- [11] S.Messelodi and C. Modena 1999. Automatic Identification and Skew Estimation of Text Lines in Real Scene Images. Pattern Recognition, 32(5):791-810.
- [12] R.Milewski and V.Govindaraju 2006. Extraction of Handwritten Text from Carbon Copy Medical Form Images. Proceedings of Seventh International Workshop Document Analysis Systems.
- [13] W.Niblack 1986. An introduction to Digital Image Processing. Prentice Hall.
- [14] Y.F.Pan, Xinwen.H, Cheng Lin L 2009. Text Localization in Natural Scene Images based on Conditional Random Field.
- [15] J.Sauvola, T.Seppanan, S.Haapakoski, M.Pietikinen. Adaptive Document Binarization. Proceedings of Fourth International Conference on Document Analysis and Recognition
- [16] K.Sobottka, H.Bunke, H.Kronenberg 1999. Identification of Text on Colored Book and Journal Covers ICDAR.
- [17] Wolf.C, Jolion J.M, Chassaing.F. Text Localization, Enhancement and Binarization in Multimedia Documents.Proceedings of International Conference on Pattern Recognition Vol.4.,Canada.
- [18] Wu.X. 1996. YIQ Vector Quantization in a New Color Palette Architecture. IEEE Transactions on Image Processing, Vol.5, No.2. 321-329
- [19] Q.Ye, Q.Huang, W.Gao,D.Zhoa 2005. Fast and Robust Text Detection in Images and Video Frames. Image and Vision Computing 23. 565-576
- [20] D.Q.Zhang 2004. Learning to detect scene text using a higher-order MRF with Belief Propagation 27-02.