Camera based Assistive Text Reading System using Gradient and Stroke Orientation for Blind Person

Kalai selvi U
PG Student, Dept. of Electrical and Electronics
Saveetha Engineering College,Chennai, India

Anish Kumar J
Assistant Professor, Dept. of Electrical and Electronics
Saveetha Engineering College,Chennai,India

Abstract— In this paper, we propose an accurate and robust method for detecting texts in object containing images. Stationary wavelet transform is applied to identify the text regions from the image. Text regions and non-text regions are separated by the connected component analysis. The non-text regions and other region in the image are removed. Character candidates are grouped into text candidates using Edge Distribution & stroke and gradient orientation maps. Canny edge detection algorithm is proposed to extract the feature from the image. Text characters are recognized using naïve's bayes classifier. The recognized text in the images are classified and converted to speech using text to speech converter using MATLAB toolbox and produced as audio to the blind people.

Keywords- Stationary wavelet transform, Morphological operator, Edge Distribution, Stroke and gradient orientation, canny edge detection.

I. INTRODUCTION

Disability of visual text reading has a huge impact on the quality of life for visually disabled people. This paper presents framework to extract text regions from images with complex background and recognize the text. Detecting text in natural image is an important in blind person for their daily life. In this paper, assistive text reading system is proposed which is useful for visually impaired person to read text from the image. Fast and robust image processing techniques are used to improve the quality of the image. This framework consists of three module 1.pre-processing 2.feature extraction 3.classification.The main objective of this project is to overcome the complexity of text reading with the complex background. In this project, stationary wavelet transform is used for time-frequency domain analysis which is applied for detect the text from image. This text image consists of four components which are used for detecting the edges.This paper is mainly focused to detecting and recognizing text from the natural image. Morphological operation can be performed to expand the text from image by structuring element. Text regions are identified by connected component analysis method. Then non-text regions are removed. Gradient and stroke orientation method is used to find the magnitude and direction of each pixel which is used to reduce the dimensionality of the image. Edge of each character can be determined based on this gradient and stroke orientation method.

Canny edge detection algorithm is proposed to extract the feature from the image. These extracted features can be compared with the template which is already created by the Naves bayes classifier. The algorithm used for text recognition is canny edge detection algorithm. This algorithm is used to filtering the noise in the natural image before detect any edges from the image. Then it calculates edge direction from the original image in the x and y direction. After finding the edge direction, the direction that can be traced in an image is related with the edge direction which is already known.Once the edge direction is known, we can apply non-maximum suppression which is used to suppress any pixel in the original image which is not considered to be an edge. The detected text in the images are converted to speech using optical character recognition (OCR) and produced as audio to the blind people.

II. RELATED WORK

Various approaches have been developed in the past for text detection and recognition of text from the image. Xilin Chen, Jie Yang, Jing Zhang, and Alex Waibel, in their work proposed automatic detection and recognition of signs. The efficient and robust method can be used to detect and recognize signs from the original image. The edge-based method is used for coarse detection accompanied by a multi-resolution scheme for different sign sizes. By combining the layout analysis and affine rectification, we obtain a markedly improved detection rate. An intensity-based OCR method is employed for sign recognition, where we apply a localized normalization method to enhance feature extraction, and use feature space.In an early work, B. Epshtein, E. Ofek, and Y. Wexler, a novel image operator that seeks to find the value of stroke width for each image pixel, and demonstrate its use on the task of text detection in natural images. The suggested operator is local and data dependent, which makes it fast and robust enough to eliminate the need for multi-scale computation or scanning windows. Extensive testing shows that the suggested scheme outperforms the latest published algorithms. Its simplicity allows the algorithm to detect texts in many fonts and languages. Kwang In Kim, Keechul Jung, and Jin Hyung Kim, proposed a texture-based method for detecting texts in images was presented. The system analyzes the textural properties of texts in images using an SVM and locates the text regions by operating CAMSHIFT on the texture classification results. This method can facilitate fast text detection, even though it does not assume the type of media or the color and textural properties of texts and is relatively insensitive to image resolution. It also works well in extracting texts from complex and textured backgrounds and was found to produce a better performance than some other techniques. However, the proposed texture classifier did encounter problems classifying very small text or text with a low contrast.A.Dutta, U. Pal, P. Shivakumara, A. Ganguli, A. Bandyopadhya and C. L. Tan, developed a simple but novel method based on gradient and edge for text detection. It is also shown that the method performs well in complex background and low resolution. Based on experimental results, it is revealed that the proposed method outperforms several existing methods in terms of detection rate, false positive rate and misdetection rate. In our method the false alarm rate is significantly lower than these existing methods. Problem is when the texture of the non-text portion coincides with the texture of the text in the images.

Hung-Hsin Chang and Hong Yan a method to proposed an extract strokes from an off-line handwritten Chinese character. We have developed four new techniques: 1) a new thinning algorithm based on Euclidean distance transformation and gradient oriented tracing, 2) a new line approximation method based on curvature segmentation, 3) artifact removal strategies based on geometrical analysis, and 4) stroke segmentation rules based on splitting, merging and directional analysis. Using these techniques, we can extract and trace the strokes in an off-line handwritten Chinese character accurately and efficiently.Comparing to the previous work, this paper can overcome the disadvantages of previous method.

III. PROPOSED METHODOLOGY

The block diagram of the pre-processing is illustrated in Figure 1.

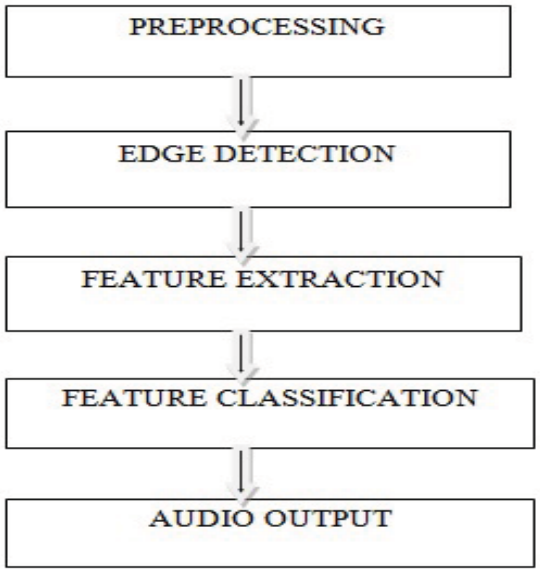

Fig.1 Block diagram of proposed system

IV. PREPROCESSING

Input image is given to the system which is resized the desired pixel rate of 256x256. This captured image is a color image which is converted into binary image for further processing of the image. Filtering can be applied for smoothing of the original image which is used to reduce the noise in the original image. This method is significant for segmenting the images. This paper proposed an average filter for smoothing process. Each pixel from the original image can be re-placed by average pixels in the 3x3 window which is created by averaging filter. Histogram equalization technique is proposed for image enhancement. The quality of the image is improved by adjusting the pixel intensity of the image.

A.1 Stationary wavelet transform

Haar wavelet transform is a stationary wavelet transform which is used to decompose the original image into four components i.e. approximate image, horizontal, vertical and diagonal image. This transform is used for converting time-frequency domain analysis which is applied for detecting edge from the images. A single image can be decomposed into multiple components using wavelet transform. Edge of each component is determined by applying sobel operator in the image and these diagonal, vertical and horizontal images are combined to produce an original image.

A.2 Morphological operation

Morphological represents the shape of text from the image i.e, Transformation of pixels in the image which is applied only in binary images. Once the edges of image is known, perform Morphological operation to expand the text using structuring element such as circle, square, rectangle in the image which can fill the holes between the image and disjoint regions are connected. The dilation operation is performed by placing structuring element on the original image. This morphological operation represents the shape of regions and will affect the structure of text image. The dilation adds pixels to the boundaries of objects in an image.

V. EDGE DETECTION

B.1 Canny edge detector

Edge detection is the robust technique to determine the edge points in the text image. The edge detection method is particularly used to find the sharp brightness changes in intensity values. Edge detection can significantly degrade the amount of information and filter unwanted information.Canny edge detector is the optimal edge detector which detects the edges from natural image in a robust manner.

The edge can be detected by perform following steps:
STEP 1: Convert color image into gray scale image
STEP 2: To reduce noise in original image
STEP 3: Calculate gradient magnitude and direction
STEP 4: Apply non maximum suppression
STEP 5: Apply Hysteresis Thresholding
After finding the edges in the image, the features of each character can identify from the image and then compared with the template which is already created by naïve bayes classifier. This classier is used to classify each character from the image. This algorithm first reduces the noise from the image.

Then Gaussian filter is used to produce the single mask in the original image. After elimination of noise, edge strength and direction is calculated by taking gradient of image. After edge direction is determined, non maximum suppression is applied to eliminate the edge pixel which is not considered to be an edge. Then Hysteresis Thresholding is applied to eliminate streaking from the original image.

VI. FEATURE EXTRACTION

C.1 Gradient and stroke orientation

Gradient image is defined as a sudden change in intensity with respect to the direction of the image. The gradient image can be computed in the x and y direction, which is used to extract the text information from the original image. This gradient orientation is used to reduce the dimensionality of the image. This orientation

method is used to calculate both direction and magnitude of image. Partial derivation is applied to each image pixel in the original image and produces the gradient image. The gradient direction and magnitude can be determined by using this equation.Canny edge detection is applied for extracting the edge of a text image which has a large magnitude.

Fig (a) Fig (b) Fig (c)

Fig .2 a) Grey scale image b) Gradient direction and magnitude image c) stroke orientation image

Each character has a multiple strokes in different orientation which is used to determine the edges in the images. The stroke orientation defines a local structure of the character. Stroke width is calculated by intensity variation in the horizontal scan line around the edge pixels. The stroke orientation is perpendicular to the gradient orientation along its gradient direction. The stroke orientation is calculated by combining gradient magnitude and direction.

C.2 Text features

Text features are used to recognize the text from the input image. These features are significant for training the image and stored in database. The Features are:

1. Major axis length: This specifies the length of the major axis of the ellipse.
2. Minor axis length: This specifies the length of the minor axis of the ellipse.
3. Eccentricity: This specifies ratio of the distance between ellipse and its major axis length.
4. Orientation: This specifies the angle between x axis and the major axis of the ellipse.
5. Convex area: This specifies the number of pixel in convex image.
6. Equiv-diameter: This specifies the diameter of the circle as same area as the region.
7. Solidity: This specifies the proportion of the pixel in the convex hull that is also in the region.
8. Extent: Ratio of pixels in the region to the pixels in the bounding box.
9. Perimeter: This specifies the number of pixel in the boundary.
10. Centroid: This specifies the centre of mass of the object.

VII. CLASSIFICATION

D.1 Naïve bayes classifier

Naïve bayes algorithm, which is used for classifies each character from the input image. The naïve bayes algorithm is based on bayes rule with strong independence classification i.e, features are unrelated to one another. Naïve bayes classifier can be trained in a supervised learning setting which gather a general features from the input image. This bayes model is based on the method of maximum likelihood which is used to extract the significant parameters from the image.

This classifier is trained by using training and testing sets. The template is developed by trained each character and compare this template image with the original image. Kernel distribution is used to calculate the kernel density for each feature within the class. Then it produces an output image. The advantage of this classifier is to required only small amount of training data.Bayes Rule:

$$P(H/E) = P(H/E) \times P(E)/P(H)$$

P (H) = probability of an event before the evidence is observed.
P (H/E) =Probability of an event after the evidence is observed.

D.2 Text to speech converter

Text to speech converter is used for convert the text in to speech. This system can recognize and classify each character from the original image. The text from the image was recognized based on the segmented image.

VIII. EXPERIMENTAL RESULTS

Fig 3(a) shows the input image which is processed in the MATLAB image processing toolbox. Fig (b) shows the original color image is transformed into grey scale image for further processing of images. Grey image is calculated by taking average of each component in the color image.

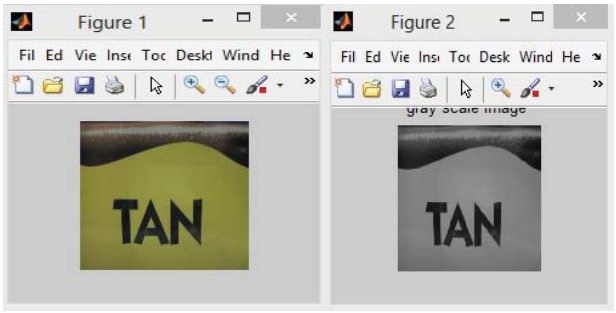

Fig (a) Fig (b)

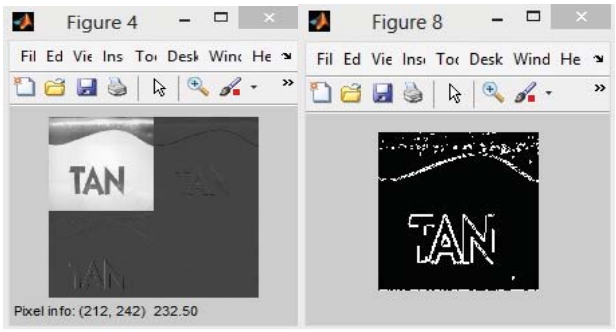

Fig (c) Fig(d)

Fig.3 a) original image b) grey scale image c) wavelet components d) combined image of wavelet components image.

The above figure 3(c) shows the wavelet based component i.e, diagonal, vertical, horizontal images. By applying wavelet transform which translate time to frequency domain of an image. Then fig 3(d) shows the combined image of wavelet components. Using this transform, we can find the edges of wavelet components by sobel operator.

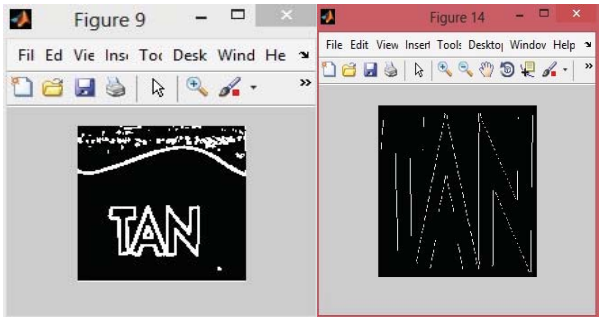

Fig (e) Fig (f)

Fig. 4 e) Morphological image f) gradients and stroke image.

The above figure 4(e) shows the morphological image which denotes the shape of the text. Dilation operation was performed to expand the text using the structuring element. This structuring element compares

neighboring pixel in the original image and produces the output image which is based on the shape of structuring element.Figure 4(d) shows the gradient and stroke image .This is the technique used for determine the edge of an image. After finding the edge, feature of an each text is extracted and classified for recognizing the text. Finally it is produced the output image using text to speech converter in the matlab.

IX. CONCLUSION

This paper proposed an accurate and robust method for detecting texts in object containing images. First input image is resized and converted into grey scale image. Then image enhancement techniques are used to improve quality of an image. By using average filter, noises in the images are removed and wavelet transform is applied by convert time domain into frequency domain analysis for simplifying the process of image. Then text and non-text regions are separated by connected component analysis. Then edges of a text image can be determined by gradient stroke distribution. After finding the edge, feature of each text is calculated for recognizing the text in the image. Then each character in the image was classified using naïve bayes classifier with kernel distribution which is used to predict the output image from the original image. Finally it is produced the speech by using text to speech converter in the matlab.Our future is to enhanced this techniques for detecting text in different structures, fonts styles etc.

REFERENCES

[1] X. Chen and A. L. Yuille, "Detecting and reading text in natural scenes,"in Proc. Comput. Vision Pattern Recognit. 2004, vol. 2, pp. II-366–II-373.
[2] P. Shivakumara, T. Phan, and C. L. Tan, "A gradient difference based technique for video text detection," in Proc. Int. Conf. Document Anal. Recognit. 2009, pp. 66–70. J. Zhang and R. Kasturi, "Extraction of text objects in video documents recent progress," in Proc. IAPR Workshop Document Anal. Syst., 2008, pp. 5–17.
[3] C. Yi and Y. Tian, "Assistive text reading from complex background for blind persons," in Proc. Int. Workshop Camera-Based Document Anal. Recognit. 2011, vol. LNCS-7139, pp. 15–28.
[4] C. Yi and Y. Tian, "Text string detection from natural scenes by structure based partition and grouping," IEEE Trans. Image Process., vol. 20, no. 9, pp. 2594–2605, Sep. 2011
[5] J.F. Canny, 'A Computational Approach to Edge Detection', IEEE Trans. Patt. Anal. Mach. Intell., PAMI-8 no. 6, Nov. 1986, pp. 679-698.
[6] B. Epshtein, E. Ofek and Y. Wexler. Detecting Text in Natural Scenes with Stroke Width Transform. In Proc. CVPR 2010.
[7] J. Liang, D. Doermann and H. Li. Camera-based Analysis of Text and Documents: A Survey. International Journal on Document Analysis and Recognition, 7(2), 2005, pp. 84–104.
[8] T. Q. Phan, P. Shivakumara, B. Su and C. L. Tan. A Gradient Vector Flow-Based Method for Video Character Segmentation. In Proc. ICDAR 2011.
[9] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. In Proc. CVPR 2005.
[10] ICDAR 2011 Dataset. http://algoval.essex.ac.uk/icdar/Datasets.html
[11] R. C. Gonzalez and R. E.Woods, Digital Image Processing, Third Edition, 2008
[12] A. Y. Ng and M. I. Jordan. On discriminative vs. generative classifiers: A comparison of logistic regression and naïve bayes. In Neural Information Processing Systems, 2002.