Middle-East Journal of Scientific Research 24 (S2): 405-409, 2016
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2016.24.S2.27

Printed Text Recognition and Speech Output for Blind Users by Speech Synthesizer

A. Shamim Banu, M. Baritha Begum, V. Kavitha, M. Padmaa and S. Priyadharshiini

Saranathan College of Engineering, Trichy, India

Abstract: In this proposed work a camera based assistive text recognition framework to aid visually challenged people for reading text label is demonstrated. The fundamental part is image segmentation which is achieved using bounding box algorithm. Calculating geometrical properties plays an essential role in recognizing the character. The segmentation process proposed in this work is suitable for the various kinds of data images including printed text. The geometrical patterns which are highlighted after segmentation is found to exhibit clear boundary which inturn provide high accuracy rate in character recognition. Extracted features vectors from gaborfiltering are passed through Neural Network classifier. Finally we get the recognized word and this is given to speech analyzer and provide audio output for the given image.

Key words: Image segmentation · Bounding box algorithm · Character recognition · Gabor filter

INTRODUCTION

Reading is obviously essential in today's society. Printed text is everywhere used in the form of reports, receipts, bank statements, product packages and instructions on medicine bottles etc. The method used here will help the visually impaired people to read the printed labels and product packages independently. There are few portable systems available but they are not suitable for identifying product labels.In the existing method, region of interest (ROI) is detected in the video from the extracted ROI, the text localization and recognition are conducted to acquire text information to automatically localize the text regions from the object ROI, using text localization algorithm features such as strokeorientations and distributions of edge pixels is done using Adaboost model. Text characters in the localized text regions are then binarized and recognized by optical character recognition software[1].In this proposed work blind persons can read text using camera based assistive text reading framework which track the object of interest within the camera view and extract only the printed text information from the object. To perform the character recognition, our application has to go through four important steps. The first step is pre-processing i.e: conversion of color image into binary image. The second step is segmentation, where each characters have been isolated using bounding box algorithm. Then next step involvesfeature extraction using Gabor filter. Since each characters can be written in a number of ways it may differ in shape and properties, such as Tilt, stroke, orientation, overall shape, so Gabor filter is designed for extracting 24 feature vectors(consists of mean and standard deviation) and then additionally region features of 40 vector numbers are combined to form 64 feature vectors (region features 40 vector size consists of area, eccentricity, perimeter, minor axis length, major axis length, orientation, centroid, equivalent diameter, extent). A various types of fonts can be used in a text that can be easily recognized and read by human but it is difficult for a machine to identify, so training and testing of character is done using artificial neural network. A feature vector of numbers will be generated which acts as the input for the artificial neural network(ANN). Then therecognized text from ANN is written into document(MS word).

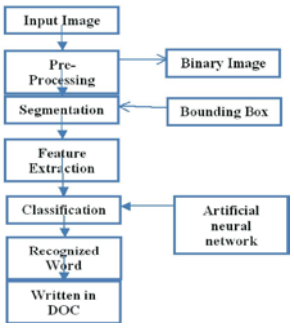

Fig. 1: Overall Block Diagram for Character Recognition

Corresponding Author: A. Shamim Banu, Saranathan College of Engineering, Trichy, India.

Steps in Character Recognition: The basic steps involved in character recognition are:

- Image capturing&Pre-processing
- Segmentation
- Feature Extraction
- Training and recognition
- Recognized text to document

Character Recognition Steps

Image Capturing & Pre-processing: An image captured is given as the input image which is to be in. JPEG format. Every image must undergo Pre-Processing where a set of operations are to be performed where the noise from the text is removed using Gaussian filtering and then binarization is done. The role of Pre-processing isto extract the desired text from the background.

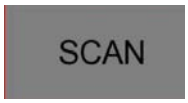

Fig. 2: Input Image

The two-dimensional digital Gaussian filter can be expressed as:

$$G(x,y) = \frac{1}{\sqrt{2\pi\sigma}} e - \frac{x^z + y^z}{z\sigma^z}$$

where σ^2 is the variance of Gaussian filter and the size of the filter kernel l(-l ≤ x, y ≤ l) is often determined by omitting values lower than five percent of the maximum value of the kernel [2].

Steps Involved in Pre-Processing:

- The input image is to be stored in JPEG format which will be in RGB color image. It is to be converted into grayscale image .It represents an image of matrix where every element has a value corresponding to the bright/dark portion of the image and the pixel value lies between 0 to 255.
- Binarization process converts a gray scale image into a binary image using global thresholding technique. This image format also stores an image as a matrix but can only color a pixel as black or white. It assigns (0)for black and (1) for white.
- Removal of background pixels to get the desired text.

Segmentation: The boundary detection of image is done for extracting the characters separately by bounding box algorithm which detects the edge pixels and the text is ready for segmentation. In the segmentation stage an image sequence of characters is decomposed into sub-images of individual characters. In this proposed work, the pre-processed input image is segmented into isolated characters by assigning a number to each character using labeling process. This labeling gives information about number of characters in the image. The segmentation algorithm will work for the text written in black with white background as well as black background with white color text.

Fig. 3: Targeted Region

Feature Extraction: In this step, the features of the characters are extracted. For feature extraction we will use Gabor feature extraction method. It improves the accuracy of recognizing a text. A 2D Gabor filters are used which extracts the features by varing the frequency and orientation of the text. In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave and additionally region based features like area, height, width of the word are also consider as a feature. Gabor features capture a number of salient visual properties, including spatial localization, orientation selectivity and spatial frequency selectivity. A set of Gabor filters with different frequencies and orientations may be helpful for extracting useful features from an image. Gabor filters have been widely used in pattern analysis applications. The Gabor function for the specified values of the parameters "wavelength", "orientation", "phase offset", "aspect ratio" and "bandwidth" will be calculated and displayed as an intensity map image in the output window. (Light and dark gray colors correspond to positive and negative function values, respectively.). The image in the output window has the same size as the input image.

Gabor Filtering (Pre-Processing): When we designing linear smoothing filters, the filter weights should be chosen so that the filter has a single peak, call the main lobe and symmetry in the vertical and horizontal directions. A typical pattern of weights for a 3x3 smoothing filter is:

1/16	1/8	1/16
1/8	¼	1/8
1/16	1/8	1/16

The most common example of the non-uniform kernel is the Gaussian filter, in which the coefficients are samples from a two-dimensional Gaussian function [3].

$$h(x,y) = \exp\left[\frac{-(x^2+y^2)}{2\sigma^2}\right] \quad (2)$$

The kernel coefficients diminish is size with increasing distance from the kernel's center. Larger values of produce a wider peak and, consequently, greater blurring.

There are several advantages to using a Gaussian filter:

- Two dimensional, Gaussian function is rotational symmetric. The kernel is rotationally symmetric. This means that the amount of smoothing performed by the filter will be the same in all directions.
- Large Gaussian filters can be implemented very efficiently because Gaussian functions are separable. Two-dimensional Gaussian convolution can be performed by convolving the image with a one-dimensional gaussian and then convolving the result with the same one-dimensional filter oriented orthogonal to the Gaussian used in the first stage [4].

$$h(x,y) * f(x,y) = \sum_{k=1}^{m}\sum_{l=1}^{n} h(k,l) f(x-k,y-l)$$

$$= \sum\sum e^{-\frac{(k^2+l^2)}{2\sigma^2}} f(x-k,y-l) = \sum_{k=1}^{m} e^{-\frac{k^2}{2\sigma^2}} \left\{\sum_{l=1}^{n} e^{-\frac{l^2}{2\sigma^2}} f(x-k,y-l)\right\} \quad (3)$$

The summation in brackets is the convolution of the input image f(x, y) with a vertical one-dimensional Gaussian function. The result of this summation is a two-dimensional image, blurred in the vertical dimension , that is then use as a the input to second convolution with horizontal one dimensional Gaussion.

One approach to designing Gaussian filter is to compute the mask weights directly from the discrete Gaussian distribution [5].

$$h(x,y) = k \exp\left[\frac{-(x^2+y^2)}{2\sigma^2}\right] \quad (4)$$

where *k* is normalizing constant.

Gabor Filter (Feature Extraction): A Gabor filter can be viewed as a sinusoidal plane of particular frequency and orientation, modulated by a Gaussian envelope [5]. It can be written as:

$$h(x,y) = s(x,y)g(x,y) \quad (5)$$

where s(x, y) is a complex sinusoid, known as a carrier and g(x, y) is a 2-D Gaussian shaped function, known as envelope [5]. The complex sinusoid is defined as follows,

$$s(x,y) = e^{-}j2\pi(u0x + v0y) \quad (6)$$

The 2-D Gaussian function is defined as follows,

$$g(x,y) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{z}\left(\frac{x^z}{\sigma x^z} + \frac{y^z}{\sigma y^z}\right)} \quad (7)$$

Thus the 2-D Gabor filter can be written as:

$$h(x,y) = e^{-\frac{1}{z}\left(\frac{x^z}{\sigma^z} + \frac{y^z}{\sigma^z}\right)} e^{-j2\pi(u0x+v0y)} \quad (8)$$

$$= g(x,y)\, e^{-j2\pi(u0x+v0y)} \quad (9)$$

The frequency response of the filter is:

$$H(u,v) = 2\pi\sigma x\sigma y\left[e^{-2\pi^2[(u-u0)^2\sigma x^2 + (v-v0)^2\sigma y^2]}\right] \quad (10)$$

$$= \frac{1}{2\pi\sigma x\sigma y} e^{-\frac{1}{2}\left[\frac{(u-u0)^2}{\sigma x^2} + \frac{(v-v0)^2}{\sigma y^2}\right]} \text{where, } \sigma u = \frac{1}{2\pi\sigma x}, \sigma v = \frac{1}{2\pi\sigma y}$$

This is equivalent to translating the Gaussian function by (*u0,v0*) in the frequency domain. Thus the Gabor function can be thought of as being a Gaussian function shifted in frequency to position (*u0,v0*) i.e at a distance of $\sqrt{u0^2 + v0^2}$ from the origin and at an orientation of $\tan^{-1}\frac{u0}{v0}$. In the above 2 equations, (*u0,v0*) are referred to as the Gabor filter spatial central frequency. The parameters ($\sigma x, \sigma y$) are the standard deviation of the Gaussian envelope along X and Y directions and determine the filter bandwidth.

Artificial Neural Network: The artificial neural network as defined by Schalkoff is a network composed of a number of interconnected units with each unit having input or output characteristics that implements a local computation or function. Typical neural networks operate in parallel nodes whose function is determined by the network structure, the connection strengths and the function in each node. Neural networks have the unit ability to "learn". Neural nets can be conveniently described as black-box computational methods for addressing basic Stimuli-Response processes (S-R).

On each side of the black-box (ANN) is a known set of inputs corresponding to their respective output set hence any distortion in the input of the system would employ algorithms and codes within the black-box to produce a unique output for that stimulus. Through this process the "new" output is added to the already existing set of standard neural network responses for known stimuli. It is important to note that the standard S-R pairs encoded into the artificial neural network ought to represent the stable states of the system during normal operation. The approach for "learning" by ANN's would take the form of *deterministic methods* like back-propagation and Hebbian approaches or it involves the *stochastic approach* such as genetic algorithms or simulated annealing [6].

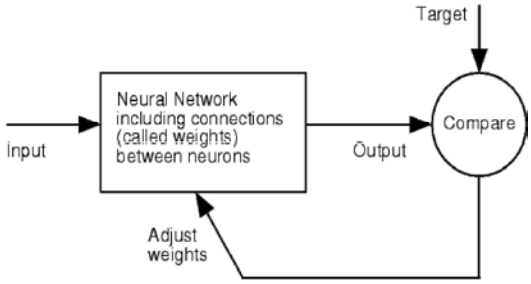

Fig. 4: General role of Neural Network.

Training: In this proposed work the training of characters (A-Z) for TIMES NEW ROMAN font with 0 boldness is done. Then it is checked with other training samples of the same character and update the weight. After that the weight matrix are save into a file, final training of the samples for that characters are performed and given as the input to the ANN classifier.

Testing: The input to ANN classifier is the combination of feature vectors and training samples. These values are tested for matching. If the given input is matched, the target number is used to display the respective character.

RESULT AND DISCUSSION

The input image captured by digital camera is a color image. Before features are extracted from an image, it is essential to pre-process the image to reduce irrelevant information or noise and to enhance the image properties that will make feature measurement easier and reliable. There may be random noise that is generated due to different factors such as dirt, dust particles so Gaussian filter is used. It can cause significant degradation in the feature extraction process which in turn may lead to higher error rates in the classification process. This noise removal is therefore essential for the system.

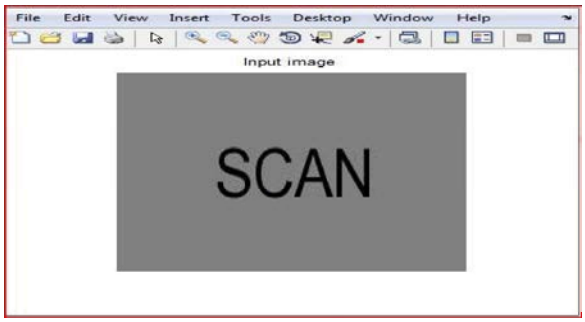

Fig. 5: Input Image

Fig. (5) refers to the input given to the proposed method.

Fig. 6: Gaussian Filtered Image

Fig. (6) refers the image is noise free after pre-processing using Gaussian filtering.

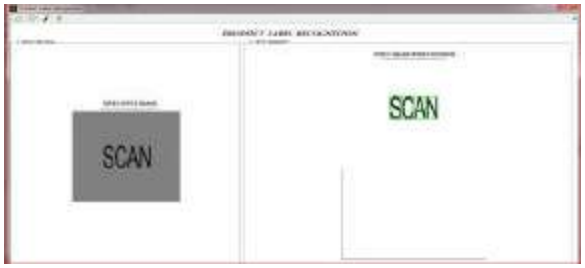

Fig. 7: Region of Characters

Fig. (7) refers the targeted characters from the input image using blob.

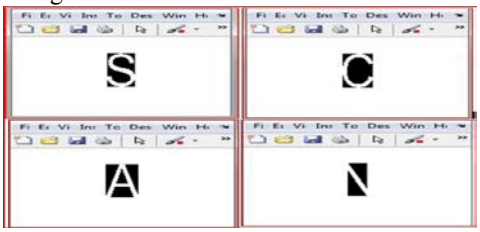

Fig. 8: Segmented Characters

Fig. (8) refers the isolated characters after using bounding box algorithm

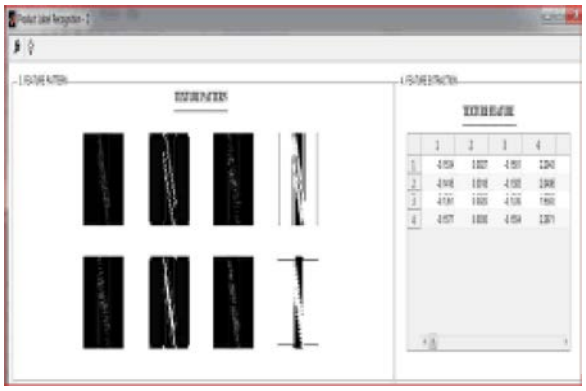

Fig. 9: Feature Extraction for the Character 'N'

Fig. (9) refers the extracted features from gabor filter for the input text 'SCAN' where 'N' is shown above. The table refers the feature vectors.

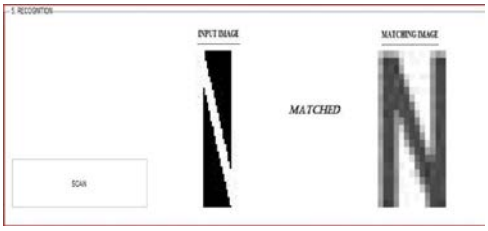

Fig. 10: Recognition of Character 'N'

Fig. (10) refers the matched character by neural network classifier.

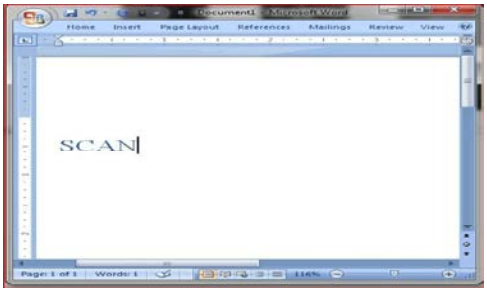

Fig. 11: Final Output

Fig (11) refers the recognized characters written to the document(MS word)

CONCLUSION

The paper gives a useful method for the recognition of printed characters to a great extent. The proposed method has been applied on different unknown characters. Neural network based method gives the accuracy. In this proposed work, with Gabor based feature extraction and neural network algorithm for feature classification is done for character recognition which give better efficiency than existing algorithm. Developed Proposed algorithm cannot be applied to recognize a cursive handwriting Recognition.

REFERENCES

1. Vasanthi, G. and Y. Ramesh Babu, 2014. "Vision Based Assistive System for Label Detection with Voice Output"International Journal of Innovative Research in Science, Engineering and Technology
2. Deng, G. and L.W. Cahill, "AnAdaptive Gaussian Filter For Noise Reduction and Edge Detection "
3. Mohammed Z. Khedher, Gheith A. Abandah and Ahmed M. Al-Khawaldeh, 2005. "Optimizing Feature Selection for Recognizing Handwritten Arabic Characters" Proceedings of World Academy of Science, Engineering and Technology, Volume 4 February 2005 ISSN 1307-6884.
4. Diego J. Romero, Leticia M. Seijas and Ana M. Ruedin, 2007. Directional Continuous Wavelet Transform Applied to Handwritten Numerals Recognition Using Neural Networks JCS&T, 7(1).
5. Plinio Moreno, Alexandre Bernardino and Jos´e Santos-Victor, 2005. "Gabor Parameter Selection for Local Feature Detection " 2nd Iberian conference on pattern recognition and image analysis, estoril, Portugaljune 7-5 2005.
6. Md Fazlul, Kader and Kaushik Deb, 2012. "NEURAL Network-based English Alphanumeric Character Recognition" International Journal of Computer Science, Engineering and Applications (IJCSEA), 2(4)..
7. Yi, C. and Y. Tian, 2011. "Text string detection from natural scenes by structure based partition and grouping," IEEE Trans. Image Process., 20(9): 2594-2605.
8. Jun Baba and Akihiro Yamamoto, "Text Localization in Scene Images by Using Character Energy and Maximally Stable Extremal Regions" IEEE Trans Image Process, 9: II-252-303.