PNN Based Character Recognition in Natural Scene Images

Shivananda V. Seeri, J.D. Pujari and P.S. Hiremath

Abstract--- The objective of this paper is to propose a novel approach to recognize Characters (Digits, Upper case letters and Lower case letters) of text spotted in natural scene image with cluttered backgrounds. The projection profile method is used to segment localized text component into individual characters, which is then submitted to character recognition process. In the proposed methodology, the Moment Invariants (MI) features are extracted from the Contourlet transformed image of segmented input characters. The PNN classifier is used for character recognition and its performance is compared with KNN classifier. The testing has been done using images taken from ICDAR dataset and personal dataset. The performance is measured in terms of recognition accuracy. The experimental results demonstrated the efficiency of the proposed method.

Keywords--- Natural scene images, Contourlet transform, character recognition, PNN classifier, Moment Invariants.

I. INTRODUCTION

TEXT information in a scene image is the key clue for scene understanding. However, the clutter background and variations, which are intrinsic in scene images, make the natural scene character recognition more difficult task. The text present in natural scene image, provide valuable content to understand the image and plays an important role in many applications such as object recognition, assistive navigation, image-based search, etc., and hence become applicable tool in wearable devices for visually challenged persons, foreign tourists and robotics. However, any method of extracting text from natural scene images must deal with challenges such as, wide variations in fonts, sizes, colors, and deformations of text characters in natural scenes, surfaces, such as signage boards, walls, cars, and objects, illuminations, view angles, complex backgrounds and difficulty in defining a general text structure model.

Due to growing significance of text recognition in natural scene images, much study has been carried out on the extraction of text and recognition in images. The various methods have been suggested for the text localization, character extraction and recognition. These methods are based on several concepts such as edge detection algorithm, morphological processing, wavelet transform, artificial neural network, histogram techniques and so on.

Text consists of set of words which are formed by sequence of characters of varying size and styles. It is hard to develop a common text descriptor for any natural language. Although many optical character recognition (OCR) systems were well developed to identify text characters in document images, methods of text recognition in scene images need to be more refined and robust. In general, a natural scene image may contain multilingual text which compounds further difficulty in character recognition. In this paper, a novel approach for character recognition in natural scene images is proposed by considering English text only.

The text extraction task from natural scene image comprises of three phases: text detection in scene image, segmentation of text regions, and extraction of characters from text regions. The method for text localization and character extraction of the segmented text has been proposed in [08]. In this paper, the objective is to perform character recognition of the so extracted text from natural scene images. Only the English language text is considered for the study. A new representative and discriminative text descriptor is designed using moment invariants features [07], extracted from contourlet transformed scene image, to model character appearances and structures. Each English character (upper/lower alphabets and digits) is a class. The probabilistic neural network (PNN) classifier [10] is trained using training samples to predict correct class label of the extracted candidate character from natural scene images.

The two types of text that appear in an image are depicted in the Figure 1.

![Example Images Containing (a) Scene Text (b) Caption Text](image_url)

Figure 1: Example Images Containing (a) Scene Text (b) Caption Text

II. RELATED WORK

Text extraction in an image includes four phases, namely, detection, localization, extraction and recognition. The main objective of text detection and text localization is to spot all text components in a natural scene image. The diversity of fonts, variation in size, style, orientation, and complex background causes lots of challenges in designing a robust
The various methods for detection and localization of text in an image are proposed in [1-3]. The method based on GLCM features of contourlet transformed image for text localization using SVM classifier is proposed for character extraction in [8]. The algorithms for character recognition of extracted text from a scene image are proposed in [5-6].

III. PROPOSED METHODOLOGY

The proposed methodology consists of three phases: preprocessing, feature extraction, character recognition using PNN classifier. The three phases of proposed methodology are described below.

A. Preprocessing

The input image is color image of a character extracted from a natural scene image. It is converted into grayscale image, which is normalized to a standard size and then subjected to histogram equalization.

B. Feature Extraction

The contourlet transform is applied on pre-processed input image to obtain edge map. The seven invariant moments (MIs), which are well known to be invariant under translation, scaling, rotation and reflection, are extracted from the contourlet transformed image.

The MIs are measures of the pixel distribution around the centre of gravity of the character and thus allows capture of the global character shape information. The moment invariants are evaluated using central moments of the image function $f(x, y)$.

The two-dimensional moments of sampled image of size $M \times M$, that has gray function $f(x, y)$, $(x, y = 0, \ldots, M − 1)$ is computed as [9],

$$m_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} (x)^p \cdot (y)^q \cdot f(x, y) \quad (1)$$

Where $p, q = 0, 1, 2, 3, \ldots$

The moments $f(x, y)$ translated by an amount $(a, b)$, are defined as,

$$\mu_{pq} = \sum \sum (x + a)^p \cdot (y + b)^q \cdot f(x, y) \quad (2)$$

Thus the central moments, $m_{pq}$ or $\mu_{pq}$ can be computed on substituting $a = −\tilde{x}$ and $b = −\tilde{y}$ as,

$$\tilde{x} = \frac{m_{10}}{m_{00}} \quad (3)$$

$$\tilde{y} = \frac{m_{01}}{m_{00}} \quad (4)$$

$$\mu_{pq} = \sum \sum (x − \tilde{x})^p \cdot (y − \tilde{y})^q \cdot f(x, y) \quad (5)$$

The change in central moments by the application of scaling normalization is given as below.

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}} \quad (6)$$

Where $\gamma = \frac{p+q}{2} + 1$

Hu defined seven values that are computed by normalizing central moments by order three and are invariant to object scale, position, and orientation. The seven moments in terms of the central moments, are given as below [9].

$$M_1 = \eta_{20} + \eta_{02} \quad (8)$$

$$M_2 = (\eta_{30} − \eta_{12})^2 + 4\eta_{11}^2 \quad (9)$$

$$M_3 = (\eta_{30} − 3\eta_{12})^2 + (3\eta_{21} − \eta_{03})^2 \quad (10)$$

$$M_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \quad (11)$$

$$M_5 = (\eta_{30} − 3\eta_{12})(\eta_{30} + \eta_{12}) \times [(\eta_{30} + \eta_{12})^2 − 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} − \eta_{03})(\eta_{21} + \eta_{03}) \times [3(\eta_{30} + \eta_{12})^2 − (\eta_{21} + \eta_{03})^2] \quad (12)$$

$$M_6 = (\eta_{20} − \eta_{02})[\eta_{30} + \eta_{12}]^2 − (\eta_{21} + \eta_{03})^2 + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad (13)$$

$$M_7 = (3\eta_{21} − \eta_{03})(\eta_{30} + \eta_{12}) \times [(\eta_{30} + \eta_{12})^2 − 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} − 3\eta_{30})(\eta_{21} + \eta_{03}) \times [3(\eta_{30} + \eta_{12})^2 − (\eta_{21} + \eta_{03})^2] \quad (14)$$

C. Classification

The Probabilistic Neural Network (PNN) is used for classification and is described below.

Probabilistic Neural Network is feed-forward neural network used as general purpose classifier based on the Parzen PDF estimation [9]. The PNN classifier is Bayes–Parzen classifier based on estimating probability density function of the input signals. Probabilistic Neural Network (PNN) computes the probability density function (PDF) for every class, based on the training samples.

Each English character is a class. Thus, there are in all 62 classes corresponding to 62 characters including 10 digits and 52 English letters (both upper and lower cases). The $n_i$ denotes the number of sample images of ith class, $i=1,2,3,\ldots,62$.

The PDF function $g_i$ of $i^{th}$ class is:

$$g_i(x) = \max_{1 \leq k \leq n_i} \left \{ e^{-\|x−x_k\|^2/\sigma_i^2} \right \} \quad (15)$$

Output(Class) = $\max_{1 \leq i \leq 62} (g_i)$ \quad (16)

Output_Class is the recognized character as a result of PNN classification. The general architecture of the PNN is shown in the Figure 2.
The general architecture of the PNN is a multi-layered feed forward network with four different layers, namely, input layer, pattern layer, summation layer and decision layer also called as output layer [10].

The proposed methodology for character recognition is depicted in the block diagram shown in the Figure 3.

The training and testing phases of the proposed methodology are described below.

D. Training Phase

In the training phase, the input color character image is preprocessed, which is subjected to contourlet transform yielding the edge map. The moment invariants features $M_1$ to $M_7$ are computed from edge map and stored in knowledge base. This procedure is repeated for all natural scene character images in training set to obtain complete knowledge base for 62 characters (upper case letters, lower case letters and digits) of English, which is used for training PNN classifier.

E. Testing Phase

In the testing phase, the input test image is preprocessed, which is subjected to contourlet transform yielding the edge map. The moment invariants features $M_1$ to $M_7$ are computed from edge map. For the feature vector $x=(M_1,\ldots,M_7)$, the PDF value $g_i(x)$ is computed for each class $i$. The maximum of $g_i(x)$ is the class label of the given input character. This is repeated for test samples of all 62 characters.
IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

The proposed method is implemented in MATLAB R2009b and tested on Intel Core i5 processor @ 2.5 GHz 4GB RAM machine. The experimental data set contains 4654 natural scene images (including ICDAR data set), out of which 2628 are used as training images and 2026 as testing images. These images are color (RGB) images in JPEG format. The PNN classifier is trained using the training set which contains 2628 images of digits, lower case letters and uppercase letters. The proposed method is executed on sets containing digits only, lower case letters only, uppercase letters only and mixed set of all characters (digits + upper case letters + lower case letters). The results of the proposed method are depicted in the Figure 4 and the Table 1. The average accuracy of the proposed method for mixed set of characters is 79.07%, which is much higher as compared to accuracy 75.15% obtained by KNN classifier. However, it is observed that the recognition accuracies for digits and lower case letters are much higher than that for upper case letters. The effectiveness of the proposed method is evidenced by the experimental results. The sample images of English characters (digits, upper/lower case letters) used for experimentation are shown in the Figure 5.

Table 1: The Performance of the Proposed Method Using PNN and KNN Classifiers

<table>
<thead>
<tr>
<th>Characters</th>
<th>PNN</th>
<th>KNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digits</td>
<td>84.13</td>
<td>72.98</td>
</tr>
<tr>
<td>Lower case letters</td>
<td>84.04</td>
<td>74.80</td>
</tr>
<tr>
<td>Upper case letters</td>
<td>74.24</td>
<td>69.01</td>
</tr>
<tr>
<td>Mixed(digits,upper/lower case)</td>
<td>79.0</td>
<td>75.15</td>
</tr>
</tbody>
</table>

V. CONCLUSION

In this study, the novel approach is proposed to recognize extracted character in natural scene image. Only the English characters are considered for the investigation. The proposed method is based on moment invariants features obtained from contourlet transformed image and the PNN classifier. The proposed method has yielded the average accuracy of 79.07% for mixed set of characters (digits + upper case letters + lower case letters), which is significantly higher as compared to accuracy 75.15% obtained by KNN classifier. Thus, the proposed methodology has potential of being applicable in wearable devices used for computer assisted vision applications.

ACKNOWLEDGMENTS

The authors are indebted to the referees for their helpful comments and suggestions. Further, the authors are grateful to the KLE Technological University and BVB CET, Hubballi 580031, India, for the financial support under the “Capacity Building Projects” grants.

REFERENCES


AUTHORS

Shivananda V. Seeri, received the B.E. and M. Tech degrees in Computer Science and Engineering from Karnataka University Dharwad in 1992 and Visvesvaraya Technological University (VTU) Belgaum in 2001, respectively. He is pursuing Ph.D (Computer Science & Engineering) in VTU, Belgaum. He is presently working as Associate Professor, Dept. of Computer Applications (MCA), KLE...
Technological University, BVBCET Campus, Hubballi, Karnataka. He has published four papers on extraction of text from natural scene images.

**J.D. Pujari**, has obtained Ph.D. (Computer Science) in digital image processing. He has B.E and M.S degrees in Computer Science and Engineering. He has Published 5 papers in International Journals and Presented 19 papers in International conferences (Several of them are IEEE International conferences), 1 in national conference. He has been presently working as Professor and Head, Department of Information Science and Engineering, S.D.M College of Engg. & Tech., Dharwad, Karnataka.

**P.S. Hiremath**, was born in May 1952, and has obtained Ph.D. (1978) in Applied Mathematics and M.Sc. (1973) in Applied Mathematics, from Karnataka University, Dharwad, Karnataka, India. He had been in the Faculty of Mathematics and Computer Science of various institutions in India, namely, National Institute of Technology, Surathkal (1977-79), Coimbatore Institute of Technology, Coimbatore (1979-80), National Institute of Technology, Tiruchirapalli (1980-86), Karnataka University, Dharwad (1986-1993), Gulbarga University, Gulbarga (1993-2014) and has been presently working as Professor, Dept. of Computer Applications (MCA), KLE Technological University, BVBCET Campus, Hubballi. His research areas of interest are Computational Fluid Dynamics, Image Processing and Pattern Recognition and Computer Networks. He has published 168 research papers in peer reviewed International Journals and Proceedings of International Conferences.