Efficient Sanskrit Word Recognition Using Segmentation and Dual Feature Extraction Techniques

[1]Nikita Gaur, [2]U Srirekha

[1]Assistant Professor, Sridevi Women's Engineering College (SWEC), Hyderabad, India swecnikita@gmail.com

[2]Assistant Professor, Sridevi Women's Engineering College (SWEC), Hyderabad, India swecsrirekha@gmail.com

Abstract

Development of a Character recognition system for Devanagri is difficult because there are about 350 basic, modifier (“*matra”*) and compound character shapes in the script and the characters in words are topologically connected. A feature based on the combination of gradient feature and coefficients of wavelet transform is developed in this paper. In handwritten word recognition, the gradient feature represents local characteristics properly, but it is so sensitive to deformation of handwritten character. Meanwhile, wavelet transform represents the character image in multiresolution analysis and keep adequate global characteristic in different scales. In order to improve the discrimination power, we composed both local and global characteristic in a combined feature. The combination schemes are described in this paper.

Keywords – Gradient, Wavelet, Sobel, Segmentation

1. INTRODUCTION

Sanskrit is the most popular script in India. It is an ancient language and no longer spoken but written materials still exist. It is very expressive language, which has been influenced and enriched by Dravidian, Turkish, Farsi, Arabic, Portuguese and English. Thus research on Devanagari script mainly Hindi language attracts a lot of interest. It has 13 vowels and 33 consonants. They are called basic characters. Handwriting recognition has always been a challenging task in image processing and pattern recognition Handwritings of different person are different; therefore it is very difficult to recognize the handwritten characters. Hence various soft computing methods involved in other types of pattern and image recognition can as well be used for DOCR (Devanagari optical character recognition). Handwritten character recognition is a challenging problem in pattern recognition area [1]. The difficulty is mainly cause by the large variation of individual writing styles. Robust feature extraction is very important to improve the performance of handwritten character recognition system. Many features have been developed in the last decades. Srikantan et al. [2] used the gradient representation as the basis for extraction of low-level, structural and stroke-type features. They demonstrated that the gray scale is useful for character recognition, even approximation directional and coarse structural information can

yield good performance of recognition. Meng. Shi et al. [3] explained the use of gradient and curvature of the gray scale character image to improve the recognition accuracy. They presented three procedures to calculate the curvature of the equi-gray scale curves, based on curvature coefficient, bi-quadratic interpolation and gradient vector interpolation. They composed a feature vector of the gradient and the curvature by simple cross product and concatenation. The results show that the direction of gradient is necessary for shape discrimination and the composite features by the cross product achieve the higher recognition rate. Gabor feature were also applied in handwritten character recognition by many researchers recently [4][5]. They used a set of Gabor filters instead of gradient operator to convolve with the input image, the feature is extracted from the real parts of the outputs. C.L.Liu et al. [6] compared the recognition performance of Gabor feature with the gradient feature; found that Gabor feature is inferior in the large deformation character recognition task such as handwritten character. Another problem is that the Gabor feature extraction is computation cost [5]. Despite of the local features, global features are also considered in handwritten character recognition. Seong-Whan Lee [7] utilized the coefficients of wavelet transform as a feature for character recognition. They suggested that the different resolution character images characterize different structures of the character. This method actually denotes a global feature in multiresolusion analysis.

In this paper, we proposed a combined feature based on the gradient feature and coefficients of wavelet transform. Generally speaking, the gradient feature represents **local characteristic** of a character image properly, but it is sensitive to the deformation of handwritten character. Meanwhile, wavelet transform represents the character image in multiresolution analysis and keeps adequate **global characteristic** of a character image in different scales. In order to improve the discrimination of a feature, it is better to compose local and global characteristic into a combined feature. The main objective of this paper is shown by the figure which is given below -:

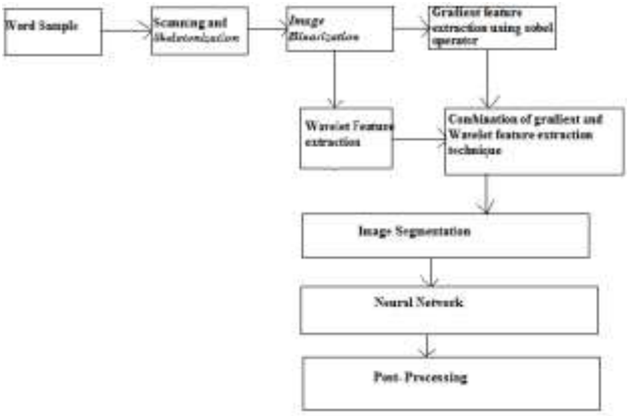

Fig 1: Block diagram of word recognition

The paper is organized as follows. Section 2 describe about neural network Section 3 gives an introduction of character modeling section 4 describes the generation of combined feature in details. Experiment results are shown in Section 5. Finally, section 6 presents the conclusion of this paper.

2. CHARACTER MODELING

2.1 Characters

The Sanskrit language consists of 53 characters (17 vowels, 36 consonants) and is written from left to right. A set of handwritten characters **is** shown in Figure 1 and 2.

अ आ इ ई उ ऊ ऋ ऌ ए ऐ ओ औ

अं अः

Fig 2 -: Set of vowels

क ख ग घ ङ च छ ज झ ञ ट ठ ड ढ ण त थ

द ध न प फ ब भ म य र ल व श ष स ह

Fig 3 -: Consonants

2.2 Scanning and Skeletonization

All character was scanned and then converted into 4096(64x64) binary pixels. The skeletonization process was used to binary pixels which were not belong to the backbone of the character, were remove and strokes were reduce to thin line.

2.3 Image Binarization

In image binarisation, the text image which is gray scale image is converted into a binary image with each pixel taking a value of 0 or 1 depending on threshold value of the image. The technique is most commonly employed for determining the threshold involve analyzing the histogram of gray scale levels in the digitized images.

I(x, y) = 0 I(x, y) <t
= 1 I(x, y)>=t

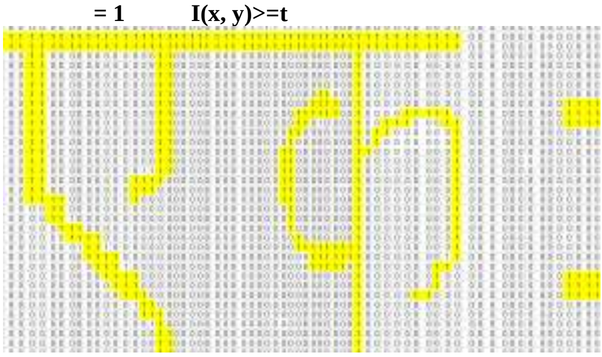

Fig 4: Binarized Image

2.4 Normalization

After skeletonization of a character, we used a normalization process, which normalized 64x64 pixel skeletonized character into 32x32 pixel characters and it was shifted to the top left corner of a window. The skeletonization and normalization processes were used for each character.

3. FEATURE EXTRACTION

The procedure of feature extraction consists of three parts, gradient calculation, wavelet transform and feature combination.

3.1 Gradient Calculation

Each gray scale image of character is normalized into 64*x* 64 size. Two gradient operators, named Sobel operator and Roberts operator[8], are used in this paper to calculate the gradient. The Sobel operators using two templates to compute the gradient components in vertical horizontal directions, respectively. The templates are shown in Fig.

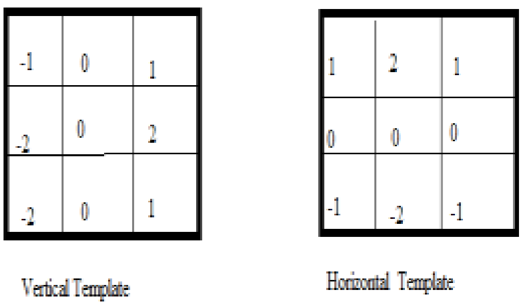

Fig 5: Sobel Operator Templates

Two gradient component at location (i, j) are calculated by:

$gv(i, j) = f(i-1, j+1) + 2f(i, j+1) + f(i+1, j+1) - f(i-1, j-1) - 2f(i, j-1) - f(i+1, j-1)$ ----------(1)

$gh(i, j) = f(i-1, j-1) + 2f(i-1, j) + f(i-1, j+1) - f(i+1, j-1) - 2f(i+1, j) - f(i+1, j+1)$ ----------(2)

The gradient strength and the direction are calculated as:

$G(i,j) = \sqrt{gv^2(i,j) + gh^2(i,j)}$ --------------------(3)

$\Theta = \arctan(Gy/Gx)$ ------------------- (4)

3.1.1 Procedure

The procedure for gradient feature extraction for handwritten character recognition is given below -:

- Capture the image of character in 30x30 pixel
- Binarization of 30x30 pixel characters into 32x32 matrixes by imposing boundary of zeros.
- Implementation of sobel operator on 32x32 matrices for gradient calculation.
- Gradient values are in a 30x30 matrix with values ranges between 0 to 2π and gradient is set to -1 where a pixel is surrounded by 8 black pixels

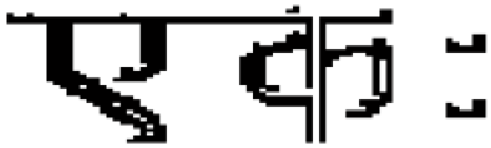

Fig 6: Gradient feature extraction

3.2 Wavelet

Wavelet transform (WT) can be regarded as a transformation that maps a signal to the multi-resolution representation. The coefficients of wavelet transform for a character image give us a scale-invariant representation in multiresolution analysis. Generally, the continuous one-dimensional wavelet transform (CWT) can decompose f (t) by a set of basis functions ψa,b (t) called wavelets

W (a, b) = $\int f(t)\, \psi^{*}_{a,b}$ (t) dt

The wavelets are generated from a single mother wavelet ψ(t) by the dilation and translation

$\psi_{a,b}$ (t) = $\frac{1}{\sqrt{a}}\left(\frac{t-b}{a}\right)$ (a>0)

Where a represents the scale factor and b represents the translation factor. In discrete wavelet transform (DWT), the scale and translation factors are discrete. In practice, the scales and translations are usually chosen as a = 2j, b = k2j. By substituting the values in above equation, we obtained:

W [j, k] = $2^{-\frac{j}{2}}$ Σ f[n][$2^{-j}n - k$]

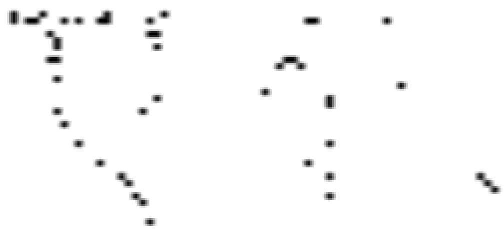

Fig 7: Noise reduced by wavelet feature extraction

3.3 Feature combination

There are two common steps must done to combine the feature extraction techniques -:

Step1 Each character image is normalized into a 64 × 64 gray scale image.

Step2 Gradient feature extraction.

- The gradient strength and directions are calculated using Sobel and Roberts's operator, respectively.
- The directions of gradient are divided into 8 ranges, such as 0^o ~ 45^o, 45^o~ 90^o, 315^o~ 360^o. Each direction range forms a direction range map. There are 8 direction range maps for a character image.
- The character image is divided into 64 blocks; one block contains 8 × 8 pixels. The gradient strengths i each block are accumulated according to the direction ranges. Each direction range map has 64 gradient strength elements.
- The gradient feature is generated from 8 range maps. The feature size is 8 × 64 = 512

Two feature combination schemes are developed and tested by us. Their brief overview is given below.

Combination Scheme 1. An image is decomposed in 3 levels using Haar wavelet in pyramid algorithm. It divided into 8 x 8 = 64 blocks, each block containing 8 x8 pixels. The directions corresponding to each of these 64 blocks are computed. Then We get 8 direction values for these blocks and further we reduce them from 8 to 4 directions in total, by considering d = d mod 4 so that the larger between two opposite directions, namely d and d mod 4, map to the smaller between them; i.e., 0 and 4 are mapped to 0 only, 1 and 5 mapped to 1 only, and so on. Thus, finally we get 4 direction codes for these blocks; but there are 64 blocks, we get a total of 64x 4 = 256 features. The coefficients of Level 3 Haar wavelet are directly used as the features without any further operation. There are 64 coefficient in total for Level 3 Haar wavelet transformations. Hence, the total feature size in Scheme 1 becomes 256 + 64 = 320.

Combination Scheme 2. Scheme 2 is similar to Scheme 1 but difference is that Scheme 2 only adds the coefficients of Level-2 Haar wavelet transformation directly along with the 320 features used in Scheme 1. Hence, Scheme 2 contain the following features inside a feature vector: the direction ranges mapped to 4 direction values: {0; 1; 2; 3}and so we get 4 x64 = 256 features, along with 16 x16 = 256 Level 2 Haar coefficients and 8x8 = 64 Level 3 Haar coefficients. Hence, the total size of feature vectors in Scheme 2 is 256 + 256 + 64 = 576.

Fig 8: Combination of gradient and wavelet feature extraction

4. SEGMENTATION

It is an operation that seeks to decompose an image for sequence of characters into sub images of individual symbols. Character segmentation is a key requirement that determines the utility of conventional Character Recognition systems. It includes line, word and character segmentation. Different method used can be classified based on the type of text and strategy being followed like recognition-based segmentation and cut classification method. After scanning the document, the document image is subjected to pre-processing for background noise elimination and skew correction to generate the bit map image of the text. The pre-processed image is then segmented into lines, words and characters. It is more clear by following example :

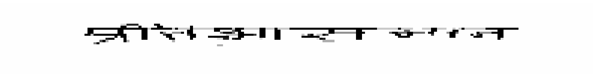

Fig 9: Original image

Firstly we remove shirorekha from each word then, we get the following :

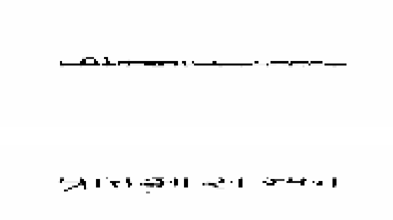

Fig 10: After removing shirorekha

Finally a character is separated from each of the word Fig. shows results explaining how a word is separated into characters.

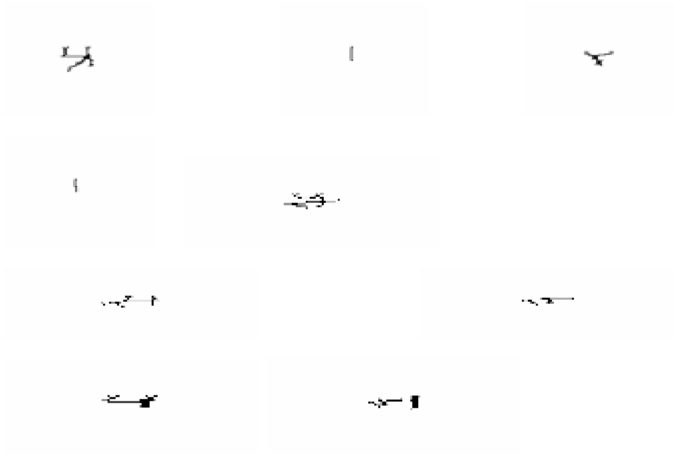

Fig 11: Separated Character

The algorithm for vertical scan is as follows.
Begin
Var: LP = left (90 degree w.r.t. vertical) neighboring pixel

Var: RP = right (90 degree w.r.t. vertical) neighboring pixel
1: for each row
2: for each column
3: x =1
4: move upward from an unvisited pixels at baseline
5: if a pixel p is unvisited & white
6: then mark p as VISITED; x =x + 1
7: else look at LP
8: if LP is WHITE
9: then move upward and goto Step 8
10: else look at RP
11: if RP is WHITE
12: then move upward and goto Step 8
13: else if (x > MedianHeight)
14: then make a vertical cut (segmented)
end

The cut separates the basic character from the vowel modifier as shown in the Fig.10. After the middle zone basic character are separated, those are sent to the recognizer for classification and editable production.

5. NEURAL NETWORK

Classification stage is the main decision making stage of the system and uses the features extracted in the previous stage to identify the text segment according to preset rules. Classification is concerned with making decisions concerning the class membership of a pattern in question. The task in any given situation is to design a decision rules that is easy to compute and will minimize the probability of misclassifications relative to the power of feature extraction scheme employed. Patterns are thus transformed by feature extraction process into points in *d* dimensional feature space. A pattern class can then be represented by a region or sub-space of the feature space. Classification then become a problem of determining the region of feature space in which an unknown pattern fall.

5.1. Neural Network Training

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

The program trains the network to recognize the characters. This network takes input-output vector pairs during training. The network trains its weight array to minimize the selected performance measure, i.e., error using back propagation algorithm.
The following are taken as inputs from the user:
a) The input pattern file
b) No. of neurons in each hidden layer
c) Value of learning rate
d) Value of momentum constant
e) Error value for convergence
The output of training program is a file which contains modified weights of different connection of the network. This file is used as the input to testing program. This file also contains the values of numbers of neurons in input layer, Hidden layers, output layer, value of learning rate and momentum factor so that used is no require to enter the values during testing.
Following parameters are used for training of Neural Networks:
No. of neurons in input layer : 7
No. of neurons in hidden layer: 12
No of epochs: 10000
Transfer Function Used for Layer 1: "Logsig"
Transfer Function Used for Layer 2: "Tansig"
Adaption Learning/Training Function: "Trainingdm"
Performance Function: MSE

5.2 Neural Network Testing

After training is complete, a test pattern is given to the neural network and the results are compared with the desired result. Difference between the two values gives the error. Percentage accuracy is found as follows:

$$\% \text{ Accuracy} = \frac{\text{No of character found correctly}}{\text{Total no of pattern}} * 100$$

6. EXPERIMENTAL RESULTS

6.1 Procedure

A complete procedure of handwritten Sanskrit word recognition is given below

- Capture the image in 64x64 pixels
- Binarization of 64x64 pixel character into 32x32 matrix by imposing boundary of zeros
- Implementation of sobel operator on 32x32 matrices for gradient calculation
- Gradient values are in 32x32 matrix with values ranges between 0 to 2pi.And gradient is set to -1 where a pixel is surrounded by 8 black pixels
- Now this 32x32 matrix having gradient values is converted into 900x1 column matrixes.
- Apply haar wavelet on 32x32 matrixes for removing noise, then combine both technique i.e. haar wavelet and gradient feature extraction.
- Apply segmentation process on 32x32 matrix and given as input to the feed forward neural network.
- This matrices can be used for training as well as simulation.
- The goal for training is set as a 2x1 matrix (e.g for character **v ,** goal =[1 0]) and for character **d** , goal =[0 1])
- By command of Matlab we will train the feed forward network.
- For simulation the command is out = sim (net, I) where the simulation result is stored in out.

The following graphs show the average number of times that a data need to be trained in order to reach a goal:

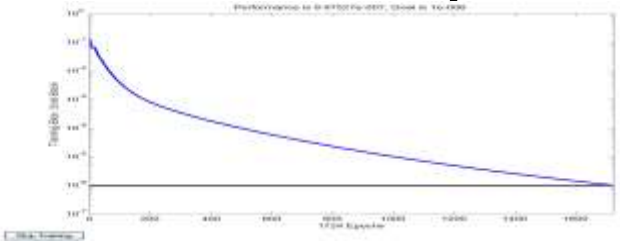

Fig12: Result of Trained Data

Above graph shows the training accuracy of system. Here, blue line denotes a training and black denote a goal or target point. In this system we trained 76 letters (including half letter and matras) .After training we compare the ideal values and practical values and the least error shows the character more visible.
Then training of system is done by using different data set or sample. And then system is tested for few of the given sample, and accuracy is measured.. An analysis of experimental result has been performed and shown in table 1.

Table 1 : Result of character recognition using Multilayer Perceptron Network

No. of Hidden Nodes(Neurons)	Learning Rate	Momentum Factor	No. of Epochs	Recognition Rate	
				Training Set	Testing Set
12	0.1	0.9	1500	100	89
26	0.1	0.9	2000	100	94
27	0.1	0.9	2500	100	96

The data set was partitioned into two parts. The first part is used for training the system and the second was for testing purpose. For each character, feature were computed and stored for training the network.

7. Conclusions and future scope

This paper shows an effort towards increasing the accuracy of handwritten Sanskrit words recognition. This work can further be extended to the word recognition of other Indian languages. For doing so, sufficient training and also a sufficient number of samples are required for the network could be trained for them. Here combination of wavelet and gradient technique has been introduced, this provides very high accuracy and less training time for handwritten Sanskrit word recognition. From the experimental result it has been shown that this combination of two feature extraction techniques is best in terms of recognition accuracy, training time and classification time.

REFERENCES

[1] Adiga, D., et al.: Improving the learnability of classifiers for sanskrit ocr corrections. Computational Sanskrit & Digital Humanities, p. 143 (2020)
[2] Avadesh, M., Goyal, N.: Optical character recognition for Sanskrit using convolution neural networks. In: 2018 IAPR 13th International Workshop on Document Analysis Systems (DAS), pp. 447452 (2018)
[3] Mridha, K., et.al.: Deep learning algorithms are used to automatically detection invasive ducal carcinoma in whole slide images. In: 2021 IEEE 6th International Conference on Computing, Communication and Automation (ICCCA), Arad, Romania, pp. 123–129 (2021)..
[4] R. Ghosh, A Recurrent Neural Network based deep learning model for offline signature verification and recognition system. Expert Systems With Applications (2020), doi: https://doi.org/10.1016/j.eswa.2020.114249.
[5] A. Dwivedi, R. Saluja and R. K. Sarvadevabhatla, "An OCR for Classical Indic Documents Containing Arbitrarily Long Words," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshop 10.1109/CVPRW50498.2020.00288.
[6]M.A. Souibgui, A.F. Biten, S. Dey, A. Foréns, Y. Kessentini, L. G omez, D. Karatzas, J. Lladós"One-shot compositional data generation for low resource handwritten text recognition" In proceedings of the IEEE/CVF winter conference on applications of computer vision (2022), pp. 935-943

[7]R. Sharma, S. Morwal, B. Agarwal, R. Chandra, M.S. Khan "A deep neural network-based model for named entity recognition for hindi language" Neural Computing and Applications, 32 (20) (2020), pp. 16191-16203

[8] Oh I.S., and C.Y. Suen, "Distance Features for Neural Network-Based Recognition of Handwritten Characters," Int'l J. Document Analysis and Recognition, Vol.1, No. 2, pp. 73-88, 1998.

[9] D.S. Yeung, "A Neural Network Recognition System for Handwritten Chinese Character Using Structure Approach," *Proceeding of the World Congress on Computational Intelligence*, **7**, pp. 4353-4358, Orlando, USA, June 1994.

[10] Dayashankar Singh, Maitreyee Dutta and Sarvpal Harpal Singh, "Comparative Analysis of Handwritten Hindi Character Recognition Technique", *IEEE International Advanced Computing Conference (IACC'09)*, March 6-7, 2009, Thaper University, Patiala.
[11] Dayashankar Singh, Maitreyee Dutta and Sarvpal H. Singh, "Neural Network Based Handwritten Hindi Character Recognition", *ACM International Conference (Compute 09)*, Jan. 9-10, 2009, Bangalore.
[12] Weipeng Zhang and Yuan Yan Tang ,"Handwritten character recognition using combined gradient and wavelet feature", IEEE International Conference on Neural Network, 2006
[13] R. Plamondon and S. N. Srihari, "On-line and off-line handwritten recognition: a comprehensive survey", IEEE Transactions on PAMI, Vol. 22(1), pp. 63–84, 2000.
[14] Huanfeng Ma David Doermann, "Adaptive Hindi OCR Using generalized Hausdorff Image comparison". \
[15] R.M.K.Sinha, "Rule based contextual post-processing for Devanagari text Recognition", Pattern Recognition", 20(5), pp. 475-485, 1987.