

RADIAL BASIS FUNCTION AND SUBSPACE APPROACH FOR PRINTED KANNADA TEXT RECOGNITION

B. VijayKumar, A. G. Ramakrishnan

Dept. of Electrical Engg.
Indian Institute of Science
Bangalore, India - 560012

ABSTRACT

Neural network based radial basis function networks (RBFN) and subspace projection approach have been employed to recognize printed Kannada characters. RBFN's are trained with wavelet features using K-means and subspace method is applied on normalized image. Use of structural features for disambiguating confused characters improved the recognition accuracy by 3% in case of subspace and by 1.6% using RBFN. Compared to subspace, a maximum recognition rate of 99.1% is achieved with RBFN using Haar wavelets and structural features.

1. INTRODUCTION

Intense research and development in optical character recognition (OCR) has led to the availability of commercial OCRs in printed Roman, Japanese, Korean, Chinese and other oriental scripts. However, the availability of such products for Indian scripts is still a rarity. The present work addresses the issues involved in designing a OCR system for printed Kannada text. Kannada, is the official language of the south Indian state of Karnataka. Modern Kannada has 48 base characters[1], called as *varnamale*. Figure 1 shows a subset of base characters from the training set. These are divided into vowels and consonants. Consonants take modified shapes when added with vowels. Vowel modifiers can appear to the right, on the top or at the bottom of the base consonant. In addition, combination of two or more characters can generate a new complex shape called a compound character. Kannada script is more complicated than English due to the presence of these compound characters. However, the concept of upper/lower case characters is absent in this script. Recognition of Kannada characters is more difficult than many other Indian scripts due to higher similarity in character shapes, a larger set of characters and higher variability across fonts in the characters belonging to the same class. There were few attempts made in the recent past for the recognition of printed Kannada characters. Probabilistic neural network classifier[2] trained with geometric, zernike and pseudo-zernike were extensively studied.



Fig. 1. Subset set of base characters from the training data normalized to 32X32

In the other approach[3] support vector machines were employed for classification using structural features extracted from subimages of a character.

In the present work, we evaluate the performance of radial basis function (RBF) networks and subspace classifier on *base characters*. Section 2 describes wavelet feature extraction method. Section 3 presents algorithmic steps in training RBF network, followed by description of subspace approach in section 4. Results and conclusions are presented in section 5, along with the structural features used to disambiguate confused character pairs.

2. DISCRETE WAVELET TRANSFORM (DWT)

Two important properties, time and frequency localization and multiresolution analysis (MRA), make the DWT a very attractive tool in image compression [4],[5]. The main advantage of MRA is that the different features of the image can be seen at different resolutions of image decomposition. Figure 2 shows the wavelet decomposition at the first level. Taking the wavelet transform of an image involves the application of a pair of lowpass (L) and highpass (H) filters called *quadrature mirror* filters. The lowpass filter corresponds to the scaling function which is the basis for the wavelet function. The highpass filter corresponds to the wavelet function. The filters H and L are applied (convolution) on the rows of the image $f(x, y)$. The resulting images $H_r f(x, y)$ (detail image) and the $L_r f(x, y)$ (approximation image) are down sampled by a factor of 2. Now filters H and L are applied on the columns of $H_r f(x, y)$ and

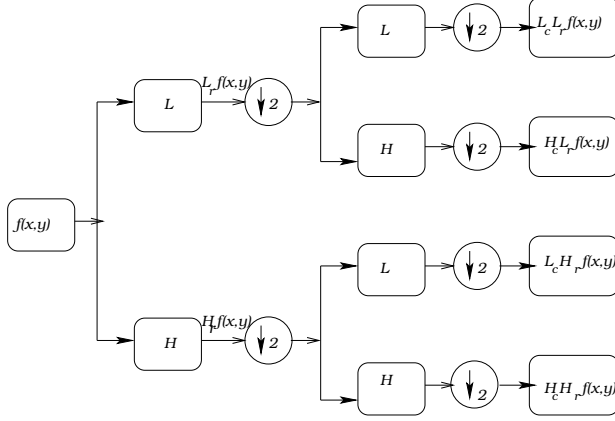


Fig. 2. Wavelet decomposition

$L_r f(x, y)$, resulting in four images namely, $H_c H_r f(x, y)$, $H_c L_r f(x, y)$, $L_c H_r f(x, y)$ and $L_c L_r f(x, y)$. And again, these images are downsampled by two. Since approximation image contains most of the energy, further decomposition is carried on $L_c L_r f(x, y)$. In our work, approximate coefficients from the normalized binary image at the second level of decomposition are used as features.

3. RADIAL BASIS FUNCTION NETWORKS

RBF network consists of three layers, namely, input layer, hidden layer and the output layer. Each node in the input layer corresponds to a component of the input vector \mathbf{x} . The second layer, the only hidden layer in the network applies non-linear transformation from input space into hidden space by employing non-linear activation function, such as Gaussian kernel. Linear nodes at the output layer correspond to the possible classes of the problem. In our method, input dimension is set to 64 and output dimension is 37 corresponding to a subset of frequently occurring characters from the base characters. Commonly used RBF is the Gaussian. A simple way to choose the number of radial basis functions is to create a hidden neuron centered on each training pattern [6]. However this method is computationally very costly and takes huge amount of memory. Therefore training patterns are first clustered into a reasonable number of groups using K-means clustering algorithm and then a neuron is assigned to each cluster center. The widths (variance) of the basis functions are set to a multiple of the average distance between the centers. The value of the multiple governs the amount of smoothing. Steps in training an RBF network [7]:

- Apply the input pattern, $\mathbf{x}_p = (x_{p1}, x_{p2}, \dots, x_{pN})^t$ to the sensory nodes of the input layer, where \mathbf{x}_p is the p th training sample or feature vector and N is the dimension of the applied input.

- Calculate the output of each hidden neuron using the Gaussian radial basis function.

$$\varphi_{pi} = \exp\left(-\frac{\|\mathbf{x}_p - \mathbf{c}_i\|}{2\sigma_i^2}\right) \quad (1)$$

where \mathbf{x}_p is the p th training sample, \mathbf{c}_i the weight vector or center of i th neuron in the hidden layer and σ_i , the width of the neuron.

- Compute the *interpolation matrix* Φ by repeating the above steps for all the training samples.

$$\Phi = \begin{bmatrix} \varphi_{11} & \varphi_{12} & \cdots & \varphi_{1N_t} \\ \varphi_{21} & \varphi_{22} & \cdots & \varphi_{2N_t} \\ \vdots & \vdots & \vdots & \vdots \\ \varphi_{N_r,1} & \varphi_{N_r,2} & \cdots & \varphi_{N_r,N_t} \end{bmatrix} \quad (2)$$

where N_t and N_r are the number of training patterns and radial basis functions, respectively.

- Find the linear weight vectors \mathbf{w} of the output layer using interpolation matrix and target vectors \mathbf{d}

$$\mathbf{w} = \Phi^{-1} \mathbf{d} \quad (3)$$

4. SUBSPACE PROJECTION

This efficiently represents higher dimensional data in lower dimensional space by projecting them onto the subspace spanned by eigen vectors corresponding to significant eigen values of covariance matrix. The steps in the computation of subspace bases[8]:

- Computation of covariance matrix:

$$\mathbf{C}_x = \frac{1}{M} \sum_{k=1}^M \mathbf{x}_k \mathbf{x}_k^t - \mathbf{m}_x \mathbf{m}_x^t \quad (4)$$

where \mathbf{m}_x is global mean vector of the ensemble of training samples, $\mathbf{x} = [x_1, x_2, \dots, x_n]^t$ with dimension 1024 and M is the size of the training set (1183 samples).

- Determination of eigen values and the corresponding eigen vectors of the covariance matrix by solving,

$$|\mathbf{C}_x - \lambda \mathbf{I}| = 0 \quad (5)$$

$$\mathbf{C}_x \mathbf{V} = \lambda \mathbf{V} \quad (6)$$

where, \mathbf{I} is identity matrix, λ is an eigen value and \mathbf{V} is the corresponding eigen vector.

- Projection of the input vectors onto the subspace by using transformation matrix \mathbf{A} formed by arranging the eigen vectors of covariance matrix in such a way

that the first row of \mathbf{A} is the eigen vector corresponding to largest eigen value and last row is the eigen vector corresponding to the smallest eigen value.

$$\mathbf{Y} = \mathbf{A}(\mathbf{x} - \mathbf{m}_X) \quad (7)$$

The transformed vectors \mathbf{Y} are used as features and the elements of the \mathbf{Y} vector are uncorrelated. By varying the dimension of transformed vectors \mathbf{Y} , recognition rate is studied.

5. RESULTS AND CONCLUSION

The data samples for the present work are collected by scanning various Kannada magazines with a resolution of 400 DPI. More than 40,000 characters are collected from these scanned images. The software is implemented on a Sun-Ultra Sparc workstation using C language and *MATLAB* package under Unix/Linux platform. The results presented are based on presegmented characters. The training set for the base characters contains 1183 samples corresponding to 37 classes and number of patterns in each class on an average is 32. The performance of subspace and RBF classifier have been evaluated on a test set containing 1453 randomly selected characters with different font styles and sizes. All the training samples are normalized to 32X32 using bilinear interpolation before feature extraction.

From the ensemble of training samples, global mean vector and covariance matrix are computed. The training samples are projected onto the subspace by using the significant eigen vectors of covariance matrix. Figure 3 shows magnitudes of first 100 eigen values. Figure 4 shows the original binary image of a character, and its reconstruction using subspace of dimension only 60. The classification performance is studied using various feature dimensions starting from 10 to 90, since first 90 eigen vectors contain most of the energy. The results using NN-classifier are shown in Figure 5. Recognition rate initially increased with increase in the size of the feature vector of subspace, but after some point, it remains constant as it is evident from Figure 5. It is observed that more than 84 percent of the total energy is stored in the first 100 eigen values. Extensive testing showed that even if more than 200 eigen values are considered, the recognition rate does not improve considerably compared to the result with a dimension of 60. These experimental results show that, 60-dimensional feature vector is best suited for recognition. Table 1 lists a group of confused characters. Structural features were used at the second level of classification to disambiguate the confused character pairs. Structural features extracted are aspect ratio, orientation of particular strokes, width of the middle zone[1] and height of the segment in the top zone from the subimages of the character. Use of structural features improved the recognition rate by around 3%. The following

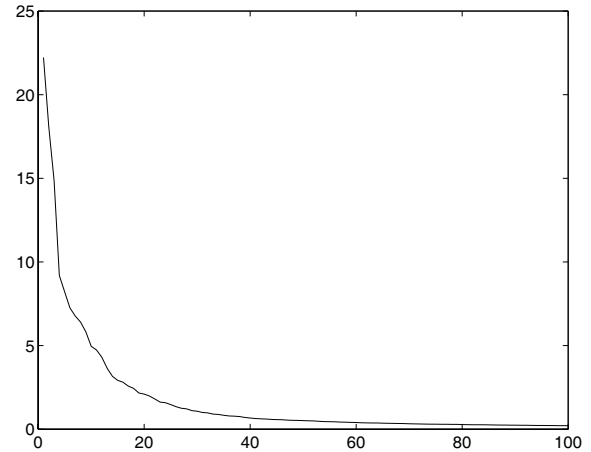


Fig. 3. Magnitudes of first 100 eigen values of the base characters normalized to 32X32

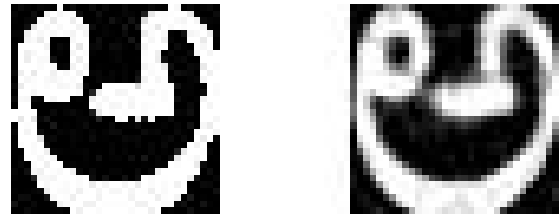


Fig. 4. Reconstruction of base character normalized to 32X32 using 60 significant eigen vectors: Original image (left) and reconstructed image (right)

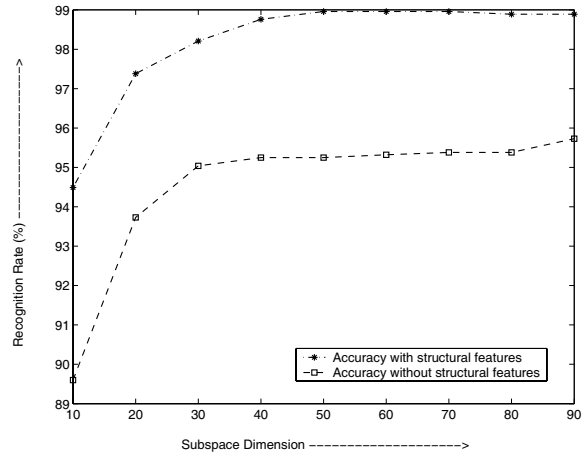


Fig. 5. Recognition rate as a function of subspace dimension

ಜ	ಪ	ಷ	ಷ	ನ	ಲ	ವ	೦
ಚ	ವ	ಪ	ವ	ವ	ಲ	ವ	ರ

Table 1. Confused character pairs using NN classifier in subspace: Top row indicates the original characters, the bottom row shows the corresponding confused characters.

itemized text shows the confused character and the structural features used to recognize it.

- ಷ: If the number of ON pixels >40 in the orientation $40^0 - 70^0$ in the lower right quarter image
- ನ: If the width of vertical projection in the bottom half of image $<75\%$ of the width of the character
- ಪ: The number of ON pixels >35 in the orientation $20^0 - 50^0$ in the upper middle region of the image
- ವ: If all the above conditions are not satisfied
- Resolving (ಠ,೦): If the width of vertical projection profile in the top and bottom half of the image is equal to its width and right matra field in middle region is empty, then it is ೦, else ಠ

Radial basis function networks are trained in supervised mode of learning employing Gaussian kernel as the activation function. The approximate coefficients in the second level decomposition of Haar wavelets are used as features with dimension of 64. On experimenting with various other wavelet families[9], Haar wavelets are found to be best for recognition. In order to reduce the computational complexity and to improve the classification speed, *K-means* clustering algorithm is used. The desired number of clusters are randomly initialized and distance between each training sample and the clusters is calculated. Nearest cluster is found out and the cluster center is updated. This procedure is repeated until no change in the cluster centers for two successive iterations. Performance of RBF is evaluated by changing the number of neurons in the hidden layer with a constant spread (variance) of 10. The number of neurons from each class in the hidden layer are selected proportionate to some percentage of total training patterns. Results with varied number of hidden neurons are shown in Figure 6. The recognition rate increased from 97.5% to 99.1% by using structural features at the second level of classification, but the improvement is only by 1.6% as against 3% using subspace classifier. Compared to subspace classifier, RBF has an overall recognition rate of 99.1% with a slight improvement of around 0.1%. Subspace technique has very good classification speed of 1.69min for 1453 characters compared to 2.51min for same set of characters using RBF networks.

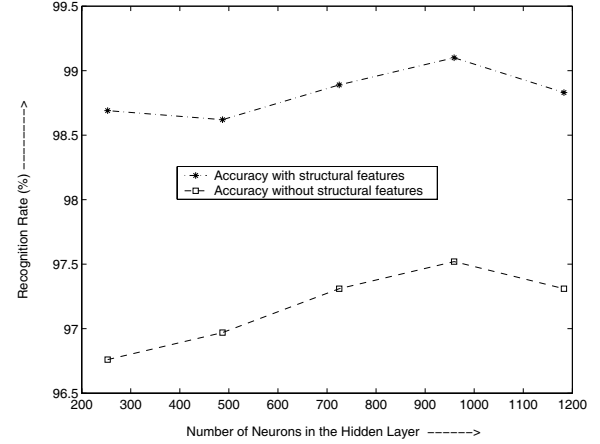


Fig. 6. Recognition rate as a function of number of neurons in the hidden layer

6. REFERENCES

- [1] B. Vijaykumar and A. G. Ramakrishnan, "Machine Recognition of Printed Kannada Text," in *Proceedings of the Fifth International Workshop on Document Analysis Systems*. 2002, pp. 37–48, Springer, Berlin.
- [2] A. Negi, B. Phanikumar, and B. K. Trinadh, "Offline printed Kannada script recognition system," In *International Workshop on Performance Evaluation Issues in Multi-lingual OCR*, Sept. 1999.
- [3] T. V. Ashwin, "A font and size independent OCR for printed Kannada using SVM," Mastersthesi, Indian Institute of Science, Bangalore, 2000.
- [4] Y. Tao, E. C. M. Lam, and Y. Y. Tang, "Feature extraction using wavelet and fractal," *Pattern Recognition Letters*, vol. 22, pp. 271–287, 2001.
- [5] D. Shen and H. S. Horace, "Discriminative wavelet shape descriptors for recognition of 2-D patterns," *Pattern Recognition*, vol. 32, pp. 151–165, 1999.
- [6] Hwang, Young-Sup, Bang, and Sung-Yang, "Recognition of unconstrained handwritten numerals by a radial basis function neural network classifier," Tech. Rep., Pohand University of Science and Technology, Korea.
- [7] S. Haykin, *Neural Networks. A comprehensive foundation*, Pearson Education Asia.
- [8] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison Wesley Publishing Co., New York, 1993.
- [9] B. Vijaykumar, "Machine Recognition of Printed Kannada Text," Mastersthesi, Indian Institute of Science, Bangalore, 2002.