

## FLD based Unconstrained Handwritten Kannada Character Recognition

Niranjan S.K.<sup>1,3</sup>, Vijaya Kumar<sup>2,3</sup>, Hemantha Kumar G<sup>4</sup>,  
and Manjunath Aradhya V N<sup>5</sup>

<sup>1</sup>*Dept of Computer Science and Engineering, Bahubali College of Engineering,  
Shravanabelgola, Hassan, Karnataka, INDIA*

<sup>2</sup>*Dept of Computer Science and Engineering, Godhavari Institute of Engg and  
Technology, Rajamundri, Andhra Pradesh, INDIA*

<sup>3</sup>*MGNIRSA, Gagan Mahal Road, Domalguda, Hyderabad, INDIA*

<sup>4</sup>*Dept of Studies in Computer Science, University of Mysore, Mysore, INDIA*

<sup>5</sup>*Dept of ISE, DSCE, Bangalore, INDIA*

*sriniranjan2007@rediffmail.com*

### Abstract

*In this paper, we propose unconstrained handwritten Kannada character recognition based on Fisher Linear Discriminant Analysis (FLD). The proposed system extracts features from well known FLD, Two dimensional FLD (2D-FLD) and Diagonal FLD. In order to classify the characters, we explore different distance measure techniques and compare their results. The proposed system is tested on unconstrained handwritten Kannada characters with pertaining to large number of character classes. The system showed effectiveness and feasibility of the proposed method.*

### 1. Introduction

Handwriting character recognition has always been a challenging and interesting task in the field of pattern recognition. Many feature extraction technique and classification algorithms have been proposed in recent years. The main business and industrial applications of character recognition in the last forty years have been in form reading, bank check reading and postal address reading. By supporting these applications, recognition capability has expanded in multiple dimensions: mode of writing, scripts, types of documents, and so on [1]. The recognizable modes of writing are machine-printed; hand printed, and script handwriting. The early recognizable scripts started with Arabic numerals and boomed to the Latin alphabets, Japanese Katakana syllabic characters, Kanji characters, and Chinese characters. The present work now being done to make Indian and Arabic scripts readable.

Several methods of recognition of Latin, Chinese, Arabic, English scripts are excellently reviewed in [2, 3, 4, 5]. In [6] a survey on feature extraction methods for character recognition is reviewed. Feature extraction method includes Template matching, Deformable templates, Unitary image transforms, Graph description, Projection histograms, Contour profiles, Zoning, Geometric moment invariants, Zernike moments, Spline curve approximation and Fourier descriptors. Different methods like neural network [7, 8], Support vector machines [9], Fuzzy logic based [10] HCR are reported for the recognition of

handwritten cursive words. Off-line Thai Handwriting recognition using Hidden Markov Model is also found in [11]. Many systems can also be seen on Indian script [12, 13, 14, 15, and 16].

Motivated by the facts that, most of the work has been done on English, Chinese, and Arabic script. Only little work on handwritten can be noticed on Indian scripts. Hence, these motivated to take up the work on totally unconstrained handwritten Kannada character recognition. Recognizing south Indian of Kannada characters is really very interesting and provides challenging in the field of document image processing due to complex orthography present in the scripts. The problem is difficult because:

1. There are about 250 basic, modified and compound character shapes.
2. The Structures of the characters have very complex orthography.
3. Kannada script is an inflectional language.

Hence in this paper, we made an attempt to recognize totally unconstrained handwritten Kannada characters based on Fisher Linear Discriminate Analysis (FLD) methods [17]. For this, we considered, FLD, 2DFLD, and Diagonal FLD based methods for feature extraction and for subsequent classification purpose, different distance measure techniques such as, Minkowski, Manhattan, Euclidean, Squared Euclidean, Mean Square Error, Angle, Correlation co-efficient, Mahalonobis between normed vector, Weighted Manhattan, Weighted SSE, Weighted angle, Canberra, Modified Manhattan, Modified SSE, Weighted Modified SSE, Weighted Modified Manhattan and Weighted Modified SSE are used.

## 2. Feature Extraction Techniques

Linear discriminant analysis [18] is usually performed to investigate differences among multivariate classes, to determine which attributes discriminate between the classes, and to determine the most parsimonious way to distinguish among classes. Similar to analysis of variance for single attribute, we could compute the within-class variance to evaluate the dispersion within class, and between-class variance to examine the differences between the classes. In this section we briefly explain FLD method for the sake of continuity and completeness. It is this success, which instigated us to investigate FLD for feature extraction and subsequent image classification. Steps involved in the feature extraction using FLD for a set of images are as follows.

Suppose that there are  $M$  training samples  $A_k$  ( $k=1, 2, \dots, M$ ), denoted by  $m$  by  $n$  matrices, which contain  $C$  classes, and the  $i^{\text{th}}$  class  $C_i$  has  $n_i$  samples. For each training character image, define the corresponding character image as follows:

1. Calculate the within class scatter matrix ( $S_w$ ) for the  $i^{\text{th}}$  class, a scatter matrix ( $S_i$ ) is calculated as the sum of the covariance matrices of the centered images in that class

$$S_i = \sum_{x \in X_i} (x - m_i)(x - m_i)^T \quad (1)$$

where  $m_i$  is the mean of the images in the class. The within class scatter matrix ( $S_w$ ) is the sum of all the scatter matrices.

$$S_w = \sum_{i=1}^c S_i$$

2. Calculate the between class scatter matrix ( $S_b$ ): It is calculated as the sum of the covariance matrices of the differences between the total mean and mean of each class.

$$S_B = \sum_{i=1}^c n_i (m_i - m)(m_i - m)^T \quad (2)$$

where  $n_i$  is the number of images in the class,  $m_i$  is the mean of the images in the class and  $m$  is the mean of all the images.

3. Solve the generalized eigenvalue problem:

Solve for the generalized eigenvectors ( $v$ ) and eigenvalues ( $\lambda$ ) of the within class and between class scatter matrices:  $S_B V = \lambda S_W V$

4. Keep first  $C-1$  eigenvectors: Sort the eigenvectors by their associate eigenvalues from high to low and keep the first  $C-1$  eigenvectors  $W$ .

5. For each sample  $Y$  in training set, extract the feature  $Z = Y^T * W$ . Then, use the nearest neighbor classifier for classification. Detailed description regarding 2D-FLD and Diagonal FLD can be seen in [17].

### 3. Experiment Results

In this section, we experimentally evaluate FLD based methods with different distance measures using FLD-based handwritten Kannada character recognition. Each experiment is repeated 25 times by varying number of projection vectors  $t$  (where  $t = 1.. 20, 25, 30, 35, 40, \text{ and } 45$ ). Since  $t$ , has a considerable impact on recognition accuracy, we chose the value that corresponds to best classification result on the image set. All of our experiments are carried out on a PC machine with P4 1.7GHz CPU and 256 MB RAM memory under Matlab 7.0 platform.

We conducted two types of experiment. In first type, Only Vowels and Consonants are considered. For experimentation we considered samples from 100 individual writers and total of 5,000 characters are considered. Here total number of classes is 50 (vowels, and consonants). Some of the sample images of handwritten Kannada vowels and consonants are shown in Fig. 1. We train the system by 75 samples and remaining samples are used during testing. Table 1 shows the recognition accuracy of different distance measures for vowels and consonants using FLD based methods. From Table 1 it is clear that for FLD based method, Angle distance measure performs better results compared to other distance measure techniques. In 2D-FLD based method, the performance of Correlation and Angle distance metric achieved good result compared to other distance measures. Whereas in Dia-FLD based method, the performance of Mahalanobis distance measure achieved good result. The combination of 2D-FLD with Angle and Correlation performs better recognition compared to other methods and distance metric.

In second type of experiment, some of the modifiers are also considered with vowels and consonants. The total number of classes considered in this experiment is 100 (vowels, consonants, and modifiers). Some of the sample images of handwritten Kannada modifiers are shown in Fig. 2. We train the system by 75 samples and remaining samples are used during testing. Table 2 shows the recognition accuracy of different distance measures for vowels, consonants and modifiers using FLD based methods. For FLD and 2D-FLD based methods, angle distance metric performs better compare to other techniques. Whereas in Dia-FLD, Euclidean performs bit better compared to other measures. Overall, Angle combined with 2D-FLD achieved better recognition rate.

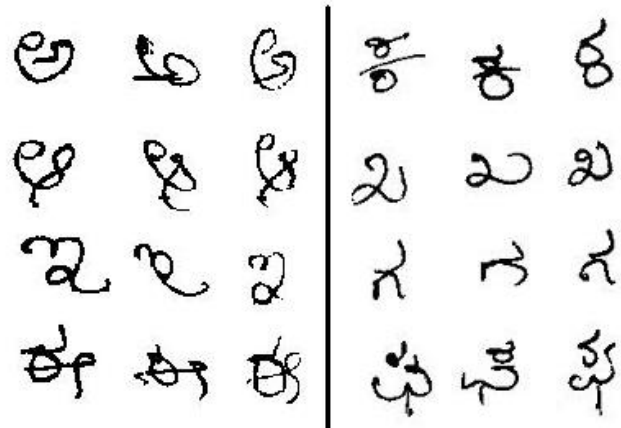


Figure 1. Sample image of Vowels and Consonants



Figure 2. Sample image of Modifiers

Table 1. Recognition accuracy of different distance Measures for vowels and consonants

Distance Measures	FLD	2D-FLD	Dia-FLD
Correlation co-efficient	67.77	<b>68.00</b>	65.55
Manhattan	63.55	62.88	60.00
Mahalanobis distance between normed vectors	67.33	65.33	<b>67.33</b>
Mahalanobis	62.22	62.44	60.11
Euclidean	66.00	66.00	64.22
Minkowski	66.00	66.00	65.00
Modified Manhattan	64.44	65.55	60.44
Modified Sq Euclidean	66.44	63.11	62.44
Mean Sq Error	66.00	67.11	62.00
Sq. Euclidean	66.00	67.22	62.00
Weight Angle	67.33	67.33	67.11
Weight Manhattan	62.88	62.11	60.88
Weight Modified Manhattan	62.22	60.00	60.77
Weight Modified SSE	65.33	64.33	64.33
Weight SSE	65.11	65.11	65.11
Canberra	51.11	48.99	46.77
Angle	<b>68.00</b>	<b>68.00</b>	66.00

Table 2. Recognition accuracy of different distance measures for vowels, consonants and modifiers

Distance Measures	FLD	2D-FLD	Dia-FLD
Correlation co-efficient	56.11	57.00	54.77
Manhattan	53.11	53.77	50.33
Mahalanobis distance between normed vectors	55.22	55.44	52.66
Mahalanobis	51.11	50.00	52.11
Euclidean	56.00	55.11	<b>55.11</b>
Minkowski	55.55	55.00	54.11
Modified Manhattan	54.11	55.11	52.99
Modified Sq Euclidean	55.66	55.66	53.77
Mean Sq Error	55.00	57.00	54.00
Sq. Euclidean	54.00	55.00	51.11
Weight Angle	56.44	54.44	52.22
Weight Manhattan	50.11	51.11	51.11
Weight Modified Manhattan	51.77	52.88	49.00
Weight Modified SSE	54.33	54.88	52.00
Weight SSE	55.44	56.00	54.00
Canberra	40.00	41.11	40.00
Angle	<b>57.00</b>	<b>58.11</b>	55.77

#### 4. Conclusion

In this paper, we addressed the problem of FLD for unconstrained handwritten Kannada character recognition. The proposed method extracts features from well known FLD, 2D-FLD and Diagonal FLD based methods. For classification purpose, we explored different distance measure techniques and tested there superiority on unconstrained handwritten Kannada characters. We conducted two types of experiments. One on vowels and consonants, another on combining modified characters. The combination of 2D-FLD with Angle and Correlation performs better recognition compared to other methods and distance metric for vowels and consonants. For the results combined with modified characters, angle combined with 2D-FLD achieved better recognition rate.

#### References

- [1] Fabien Lauera, ChingY. Suenb, Gerard Blocha, A Trainable Feature extractor for handwritten digit recognition, *Pattern Recognition*, Vol 40, 2007, pp. 1816-1824.
- [2] R. Plamondon and S. N. Srihari, "On-line and off- line handwritten character recognition: A comprehensive survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, vol. 22, no. 1, pp. 63-84.
- [3] Nafiz Arica and Fatos T. Yarman-Vural, "An Overview of Character Recognition Focused on Off-Line Handwriting", *IEEE Transactions on System.Man.Cybernetics-Part C: Applications and Reviews*, 2001, vol. 31, no. 2.
- [4] Liana M. Lorigo and Venu Govindaraju, "Offline Arabic Handwriting Recognition: A Survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, vol. 22, no. 5.
- [5] G. Nagy, "Chinese Character Recognition, A twenty five years retrospective", *Proceedings of ICPR*, 1988, pp. 109-114.
- [6] Anil.K.Jain and Torfinn Taxt, "Feature Extraction Methods for Character Recognition-A Survey: ", *Pattern Recognition*, 1996, vol. 29, no. 4, pp. 641-662.
- [7] Brijesh Verma, Michael Blumenstein, Moumita Ghosh, "A Novel Approach for structural feature extraction: Contour vs. direction: ", *Pattern Recognition letters*, 2004, vol. 25, pp. 975-988.

- [8] Michael Blumenstein, Xin Yu Liu, Brijesh Verma, "An investigation of the modified direction feature for cursive character recognition", *Pattern Recognition*, 2007, vol. 40, pp. 376-388.
- [9] Francesco Camastra, "A SVM- based cursive Character recognizer", *Pattern Recognition*, 2007, vol. 40, issue 12, pp. 3721-3727.
- [10] N. M. Noor, M Razaz and P. Manley-Cooke, "Global Geometry Extraction for Fuzzy Logic Based Handwritten character Recognition", *Proceedings of the 17<sup>th</sup> International Conference on Pattern Recognition*, 2004, vol. 2, issue. 23-26, pp. 513-516.
- [11] Roongroj, Nopsuwanchai, Alain Blem, and William F. Clocksin, "Maximization of mutual information for Offline Thai Handwriting recognition", *IEEE. Transactions on Pattern Analysis and Machine Intelligence*, 2006, vol. 28, No. 8, pp. 1347-51.
- [12] Bhattacharya U, Parui S.K, Shaw B, and Bhattacharya K, *Neural Combination of ANN and HMM for Handwritten Devnagari Numeral Recognition*, Proceedings of 10<sup>th</sup> IWFHR, 2006, pp. 613-618.
- [13] Swethalakshmi H, Jayaraman A, Chakravarthy V.S., and Sekhar C.C., *Online Handwritten Character Recognition of Devnagari and Telugu Characters using Support Vector Machines*, Proceedings of 10<sup>th</sup> IWFHR, 2006.
- [14] Pujari A.K., Naidu C.D., Sreenivasa Rao, Jinaga B.C., *An Intelligent Character Recognizer for Telugu Scripts using Multiresolution Analysis and Associative Memory*, *Image and Vision Computing*, Vol. 22, 2004, pp. 1221-1227.
- [15] Bhattacharya U, Das T.K., and Chaudhuri B.B., *A Cascaded Scheme for Recognition of Handprinted Numerals*, Proceedings of 3<sup>rd</sup> ICVGIP, 2002, pp. 137-141.
- [16] Patil P.M., and Sontakke T.R., *Rotation, Scale and translation handwritten Devnagari Numeral Character Recognition using General Fuzzy Neural Network*, *Pattern Recognition*, Vol. 40, 2007, pp. 2110-2117.
- [17] Noushath, Hemantha Kumar, and Shivakumara, *Diagonal Fisher linear discriminant analysis for efficient face recognition*, *Neurocomputing*, Vol 69, 2006, pp 1711-1716.
- [18] Fisher R A, *Ann. Eugen* , Vol 7, 1936, pp. 179-188.