Divide and Conquer Technique in Online Handwritten Kannada Character Recognition

M. Mahadeva Prasad
Dept. of Electronics
Mysore University P.G. Centre
Hassan, INDIA.
prasada9@gmail.com

M. Sukumar
Dept. of Instrumentation Technology
S. J. College of Engineering
Mysore, INDIA.
m_sukumar@rediffmail.com

A.G. Ramakrishnan
Dept. of Electrical Engineering
Indian Institute of Science
Bangalore, INDIA.
ramkiag@ee.iisc.ernet.in

ABSTRACT
The peculiar nature in which one or more consonants combine with vowels to produce a compound character in Kannada language results in a huge number of character combinations, running to tens of thousands or more. The aim of the work is therefore, to reduce the number of character combinations by employing a divide and conquer technique. In the first level of the technique, the structural and the dynamic features of online handwritten Kannada characters are exploited to segment the compound Kannada characters into 282 distinct symbols. This reduction in the number of classes overcomes the huge data collection problem and also reduces the computational complexity. In the second level, these 282 symbols are further divided into three distinct sets of stroke groups, thus further reducing the search space for the recognition engine. One or more of these stroke groups can combine to form any of the thousands of Kannada compound characters. Since the focus of this paper is the above strategy, a simple classifier has been used to validate the effectiveness of the proposed scheme in handling the difficult task of recognizing all possible character combinations of Kannada. The features extracted from the segmented stroke groups are mapped to lower dimensional space using PCA. The subspace features of distinct stroke groups are fed to the respective classifiers in an order and the output of these classifiers are combined to output the Unicode of the recognized akshara. The proposed work is an attempt made for the first time in Kannada language which considers all possible combinations of symbols, including Kannada numerals.

Categories and Subject Descriptors
I.5.2 [**Pattern Recognition**]: Design Methodology - *Classifier design and evaluation*, *pattern analysis.*

General Terms
Design, Algorithm, Experimentation, Performance.

Keywords
Online handwritten character recognition, Kannada, character segmentation, divide and conquer technique.

1. INTRODUCTION

Online handwritten character recognition (OHCR) of Indic scripts is one of the challenging tasks in the field of pattern recognition. The huge size of the character set, their complex shapes, confusion among them and variable writing styles have made it difficult to attain good recognition accuracy. In spite of these difficulties, efforts are being made to interact with machines through handwriting. In this direction, more effort has been made to realize OHCR systems of South Indian languages like Tamil [1, 2, 3] and Telugu [4, 5, 6] than Kannada. Though research work can be found in the literature for recognizing the machine printed Kannada text [7, 8, 9], to the best of our knowledge, no progress has been made to realize online handwritten Kannada character recognition (OHKCR) except the work by Kunte at al. [10]. The present work is an attempt to systematically look at the problem of developing an OHKCR system capable of recognizing all possible combinations of Kannada characters. In this paper, we address the complexity issues of Kannada characters for OHCR and propose an effective two level divide and conquer strategy (DCS). In DCS, a compound Kannada character is segmented into three stroke groups, which are recognized separately. In the literature, we can find the OHCR systems based on the stringent rule-based approach to combine the recognized strokes/radicals/ligatures for Chinese [11, 12] and Telugu [4, 5] characters, which demand high computational costs. In order to minimize the computation costs and also for a better recognition results, in this paper, a combined symbol and stroke level recognition is implemented.

In the first level of the DCS, the structural and the dynamic features of online handwritten Kannada characters (OHKC) are exploited to segment the compound Kannada characters into 282 distinct symbols. This overcomes the huge data collection problem and also reduces the computational complexity, by reducing the total number of classes to be handled by the recognition engine. In the second level, these 282 symbols are further divided into three distinct sets of stroke groups, thus

further reducing the search space. It is to be noted that combinations of one, two or more of these stroke groups can form any of the thousands of Kannada compound characters. Two sets of features extracted from the segmented stroke groups are mapped to lower dimensional space using PCA. The subspace features of distinct stroke groups are fed to the respective k-NN classifiers in an order and the output of these classifiers are combined to achieve the final recognition Unicode.

The rest of this paper is organized as follows. Section 2 describes the complexities of the Kannada script. Section 3 presents the proposed divide and conquer strategy and hence the data reduction techniques. OHKC database creation, preprocessing, feature extraction and the pattern classification scheme are explained from sections 4 to 7. Sections 8 and 9 are devoted to experimentation, conclusions and the scope for future work.

2. KANNADA SCRIPT

Kannada is the official and administrative language of Karnataka, one of the South Indian states of India. The language belongs to the group of Dravidian languages of India estimated to be over 2500 years old and has its own writing system. The Kannada script is phonetic and uses fifty characters corresponding to 50 phonemes. The glyphs of these fifty basic characters (16 vowels and 34 consonants) along with their ITRANS (Indian language TRANSliteation) are shown in Table 1. In Kannada, one or more (maximum 3) consonants (C) combine with a vowel (V) to produce a new grapheme. The consonant-vowel (CV) combinations in Kannada are formed by retaining most of the consonant glyph and attaching a glyph corresponding to the vowel modifier. The glyphs of the CV combinations for a consonant 'ಣ' (Na) are shown in Figure 1.

When two consonants combine, the second consonant changes its shape or size and is written to the right bottom of the first and is called a consonant-conjunct leading to a new grapheme. For example, a CCV combination, 'ಸ್ವ' (swa) is obtained by writing a consonant-conjunct '್ವ' corresponding to a consonant 'ವ' (va) below the vowel modified consonant 'ಸ' (sa). However, when the first consonant is 'ರ' (ra), the second consonant appears first without change in shape or size, whereas 'ರ' (ra) changes its shape and appears to the right. Thus, there are 34 consonant-conjuncts, one corresponding to each consonant. The glyphs of these consonant-conjuncts are shown in Figure 2. Some examples of CCV and CCCV combinations are shown in Table 2.

In CCCVs, the vowel modifies the first consonant and the other two form the consonant-conjuncts, written below the CV. For example, a CCCV combination, 'ತ್ಸ್ಯ' (tsya) is obtained by writing two consonant-conjuncts '್ಸ' and '್ಯ' corresponding to consonants 'ಸ ' (sa) and 'ಯ' (ya) below the vowel modified consonant 'ತ' (ta). Last but not the least, Kannada has its own numeral system, as illustrated in Table 3.

Table 1. Kannada Character Set with their ITRANS shown below the character.

Vowels	ಅ	ಆ	ಇ	ಈ	ಉ	ಊ	ಋ		ೠ	ಎ
	a	aa	i	I	u	U	Ru		RU	e
	ಏ	ಐ	ಒ	ಓ	ಔ	ಅಂ	ಅಃ			
	E	ai	o	O	ou	aM	aH			
Consonants modified by vowel 'ಅ' (a)	ಕ	ಖ	ಗ	ಘ	ಙ	ಚ	ಛ	ಜ	ಝ	ಞ
	ka	kha	ga	gha	nga	ca	cha	ja	jha	nja
	ಟ	ಠ	ಡ	ಢ	ಣ	ತ	ಥ	ದ	ಧ	ನ
	Ta	Tha	da	dha	Na	ta	tha	da	dha	na
	ಪ	ಫ	ಬ	ಭ	ಮ	ಯ	ರ	ಲ	ವ	ಶ
	pa	pha	ba	bha	ma	ya	ra	la	va	sha
	ಷ	ಸ	ಹ	ಳ						
	Sha	sa	ha	La						

ಣ	ಣಾ	ಣಿ	ಣೀ	ಣು	ಣೂ	ಣೃ	ಣೄ	ಣೆ	ಣೇ
Na	Naa	Ni	NI	Nu	NU	NRu	NRU	Ne	NE
ಣೈ	ಣೊ	ಣೋ	ಣೌ	ಣಂ	ಣಃ				
Nai	No	NO	Nou	NaM	NaH				

Figure 1. All the CV combinations of consonant 'ಣ' (Na) , along with their ITRANS.

ಕ	ಖ	ಗ	ಘ	ಙ	ಚ	ಛ	ಜ	ಝ	ಞ	ಟ	ಠ	ಡ	ಢ	ಣ	ತ	ಥ
ದ	ಧ	ನ	ಪ	ಫ	ಬ	ಭ	ಮ	ಶ		ಷ	ಸ	ಹ	ಳ	ೞ	ವ	ಱ

Figure 2. Consonant-conjuncts (or ottus, which are written below the main character) corresponding to the 34 consonants.

To summarize, a typical Kannada character can be a vowel, a consonant, a CV combination, a CCV combination, a CCCV combination or a numeral. The total number of Kannada character combinations as suggested in Table 4 is 647921, even though some of them are not phonotactically valid. So, it is almost impractical to build a recognizer that deals with such a huge number of classes. Herein lies the crux of our work - development of an efficient scheme to reduce the computational complexity of the problem at hand. We adopt two levels of divide and conquer approach to circumvent this issue, which is addressed below.

3. DIVIDE AND CONQUER THROUGH SEGMENTATION

Owing to the huge size of Kannada character combinations, considering each of them as a separate class for recognition may reduce the recognition accuracy and increase the computation cost. Also, the occurrence of various combinations of half-letters

Table 2. CCV and CCCV combinations: some examples.

CCV combination	ಕ್ಕ	ಕ್ನ	ಕ್ಯ	ಸ್ವ	ಣ್ಣ
CCCV combination	ಸ್ಮಿ	ಸ್ಕಲ	ತ್ಸ್ಯ		

Table 3. Kannada Decimal Number System.

Numerals	೦	೧	೨	೩	೪	೫	೬	೭	೮	೯
	0	1	2	3	4	5	6	7	8	9

Table 4. Calculation of possible Kannada character combinations.

Character Type	Possible Combinations
Vowels	16
CV	544
CCV	18496
CCCV	628864
Numerals	10
Total	647921

called glyphs that attach to various letters in a manner similar to diacritical marks in the Roman languages may lead to misclassification among characters as most of the characters look similar in shape, except for the glyphs attached to them. For example, the characters ರ (ra) and ಠ (Ta) vary only with respect to the presence or absence of a dot called 'bindu' in the middle of the character. Similarly, characters, ಪ (pa) and ಫ (pha) differ by a small vertical line called 'pulli' at the bottom of the middle region. Also, for data collection, asking individuals to write all the character combinations is not practical. Most importantly, this approach does not exploit the structural and the spatio-temporal information of online handwritten Kannada characters. So, we choose to divide each compound character into its component symbols and conquer the complexity. We have carefully studied the structural nature of the characters and the common methods of writing them. This analysis shows that any complex character can be obtained as a combination of basic characters, the stroke units corresponding to the vowel modifiers and the consonant-conjuncts. Hence, it is sufficient if a recognition engine is trained with these units to recognize any character combination. In order to achieve our goal, we have implemented the divide and conquer technique, in which any complex combination is segmented into three basic units and are recognized independently. The outputs of these recognizers are combined in an order to obtain the final recognition of a character. The sections to follow explain the strategy implemented to realize this.

After analyzing the glyphs of the CV combinations of a consonant 'ಣ'(Na) given in Figure 2, they are separated into three groups as shown in Table 5. It is clear from Table 5 that, by writing a stroke corresponding to the vowel-modifier to the right or to the bottom of the first group of characters, the second and the third group of characters can be obtained. Same technique is extended in grouping the CV combinations of all other consonants. Similarly, by looking at the CCV and the CCCV glyphs in Table 2, it is clear that, on writing a vowel modifier to the right and one

Table 5. Three groups of CV combinations of character 'ಣ'(Na). The first group characters have only middle and the top region strokes. Second group characters have strokes written to the right of first group of characters. Third group characters have strokes written at the bottom of the first group of characters.

Group 1	ಣ ಣಾ ಣಿ ಣೆ ಣೌ
Group 2	ಣೀ ಣು ಣೂ ಣೇ ಣೊ ಣೋ ಣಂ ಣಃ
Group 3	ಣ್ಮ ಣ್ಯಾ ಣ್ಯೈ

Table 6. Characters differing by a dot or/and pulli

Group 1	ರ	ಪ	ದ	
Group 2	ಠ	ಫ	ಧ	ಥ

or two consonant-conjuncts below the basic consonant, any CCV and CCCV combinations of that consonant can be obtained.

In addition to the above combinations, there are characters which differ by the presence of the pulli or the dot stroke or both, as shown in Table 6. It is clear from the table that, by writing a dot or/and pulli to the first group of characters, the second group of characters can be obtained. In such cases, the dot and the pulli strokes are removed from the character before sending it to the recognition engine, after setting the corresponding flags.

3.1 Segmentation

On observing the character combinations, we can notice three regions, namely, middle, top and the bottom; there exists a vertical overlap among these three regions. In addition, the order of writing a character covering these regions is middle, top, middle region vowel modifiers and then the bottom. Each region may have a single or multiple strokes. In some cases, the middle and the top regions are connected or the writing style of individuals may lead to this condition and both the regions are written using a single stroke. However, the bottom region is not connected with the middle region.

In the proposed divide and conquer scheme, the well defined rules of combining different symbols to obtain compound characters, writing methods and the dynamic pen motion of online Kannada handwriting are exploited to segment a character into three groups. They are,

1) The main unit (MU), comprising of strokes that occupy middle and the top region of the character.

2) The right auxiliary (RA) having strokes written to the right of MU and occupying the middle region.

3) The bottom auxiliary (BA) having the bottom region strokes such as conjuncts and vowel modifiers.

Figure 3 illustrates the order of writing the strokes of a handwritten character having all the three auxiliaries. After the proposed segmentation, MU consists of strokes 1 and 2, the RA consists of stroke 3 and the BA consists of stroke 4. Handwritten Kannada characters may vary by having only the MU, MU and one or more RAs or having all the three auxiliaries as shown in Table 7. The concise algorithm used to group a character into MU, RA and BA is given below.

1. To start with, the coordinates of the bounding box (BB) of each stroke are obtained.
2. The first stroke of a character is considered as the first stroke of MU. Let BB_1 be the bounding box of this stroke.
3. The BB information of the next stroke is compared with that of BB_1. Based on the position of the bounding box of the current stroke, it is appended to MU, RA or BA.
4. If the BB of the current stroke is above BB_1 and α% of its area overlaps with BB_1, the stroke is appended to the strokes of MU.
5. If the BB of the current stroke is to the right of BB_1 and even if β% of its area overlaps with BB_1, the stroke is appended to the strokes of RA.
6. If the BB of the current stroke is below BB_1 and at least γ % of its area overlaps with BB_1, the stroke is appended to the strokes of BA.
7. Steps 4 to 6 are repeated till all the strokes of a character are taken into consideration.

On performing segmentation over the existing character data, the values of α, β and γ are fixed to 75, 10 and 25 respectively. Tables 8 and 9 give an idea of the proposed spatial segmentation technique employed in segmenting a character combination into MU, RA and BA. If a character has more than one RA or BA units, they are separated and numbered accordingly.

3.2 Data Reduction

In the reduced set of character combinations, characters whose RA and BA could be isolated are not considered. Hence, the reduced set of symbols includes the first group of CV combinations of all consonants, the vowels and the vowel modified consonants in which the MU and the RA are written in a single stroke. Figure 4 shows few example character combinations in which the MU and RA are sometimes written with a single stroke due to a variation in the writing style. Let this set, along with the Kannada numerals be called Main Group (MG). Table 8 gives an idea of possible combinations of the character 'ಣ' (Na), considered under MG for training the recognition engine. However, it is necessary to train the engine with RA strokes and BA stroke groups so that the segmented RA and BA units of any complex combination can be recognized. In order to extract the required RAs and BAs, character combinations having RAs and BAs from which these stroke units can be segmented are added to the reduced set. Let this set be called Auxiliary Group (AG). Table 9 shows examples of RA and BA stroke/stroke groups extracted from character combinations written together as one unit through segmentation, for training the recognition engine. In our reduced data set, there are 225 symbols in MG and 57 symbols in

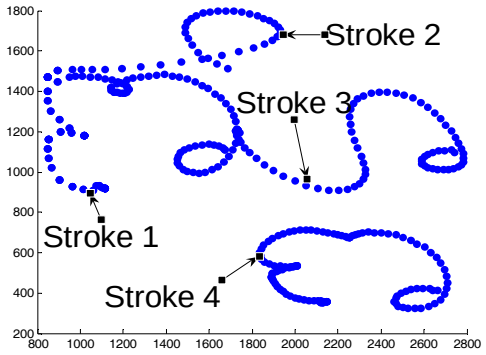

Figure 3. A handwritten character with three Auxiliaries.

Table 7. Characters grouped based on MU, RA and BA stroke groups

Characters with MU	ಅ ಔ ಕ ಚ ಣಾ ಷ
Characters with MU and RA	ಋ ಕೂ ಣೇ ಅಂ ಮಾ ಯು
Characters with MU, RA and BA	ಣ್ಣ ಮ್ಮ ಯ್ಯ

Table 8 . CV Segmented into MU, RA and BA

CV combination	Segmented constituents			
	MU	RA1	RA2	BA
ಣು	ಣ	ು	–	–
ಣಂ	ಣ	ಂ	–	–
ಣೋ	ಣ	ೊ	ೀ	–
ಣ್ಯೆ	ಣ	–	–	ೆ

Table 9 . CCV/CCCV Segmented into MU, RA and BA.

CCV/CCCV combination	Segmented Constituents			
	MU	RA	BA1	BA2
ನ್ಚ	ನ	–	ಚ	–
ಣ್ಣ	ಣ	್	ಣ	–
ಡ್ಢ	ಡ	–	ಢ	–
ನ್ತ್ಯ	ತ	–	ನ	್ಯ

AG in which 20 are meant for collecting RA strokes and the remaining 37 for collecting BA stroke groups. So, in total, there are 282 symbols in the sub-set for recognizing all character combinations and numerals of Kannada. With this, there is a huge reduction in the number of character combinations to be considered for data collection and training the recognition engine.

4. DATA COLLECTION

To our knowledge, there is no public online handwritten Kannada character database to carry out recognition experiments. So, creating one has been a part of our research work. For data collection, only the reduced set of Kannada characters is considered. The data is collected using Tablet PCs and Genius digital note pads. Each writer has written up to a maximum of five trials in his/her own writing style, without exercising any restrictions.

5. PREPROCESSING

The segmented raw stroke groups, MA, RA and BA are subjected to pre-processing which involves noise removal, re-sampling and normalization. The noise is due to the erratic pen movements, which may lead to redundant and extraneous points in the data. As the raw data is sampled in time, the number of points in a character varies based on the sampling rate of the device and the speed of writing. Also, due to the writing styles, there are variations in the size of the characters.

A moving average filter of size three, which reduces the noise and smoothes the data is used. The smoothed data is re-sampled along the arc length by linear interpolation so that all auxiliaries have 30 equal number of points. For stroke groups with multiple strokes, the number of sampled points is distributed proportional to their lengths. Let the smoothed and sampled stroke group be represented by the sequence

$$P = [p_1, p_2, ...p_{30}] \quad (1)$$

where the vectors $p_i = (x_i, y_i)^T$ denote the horizontal and vertical coordinates. The sampled stroke groups are shifted and are size normalized to get a new sequence

$$Q = [q_1, q_2, ...q_{30}]. \quad (2)$$

where the vectors $q_i = (a_i, b_i)^T$ is given by

$$a_i = (x_i - x_{\min})/(x_{\max} - x_{\min}) \quad (3)$$

$$b_i = (y_i - y_{\min})/(y_{\max} - y_{\min}) \quad (4)$$

where, $(x_{\min}, y_{\min})^T$ and $(x_{\max}, y_{\max})^T$ denote the minimum and the maximum horizontal and vertical coordinate values.

6. FEATURE EXTRACTION

The following six features are extracted from the preprocessed stroke groups to get a new sequence:

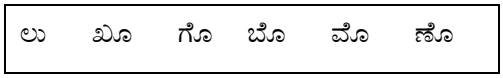

Figure 4. MU and RA, which are sometimes written in single stroke.

Table 8. Possible combinations of the character 'ಣ' (Na), considered as Main Group, for training the recognition engine.

Main Units	ಣ ಣಾ ಣಿ ಣೆ ಣೌ
Additions due to writing style	ಣು ಣೊ
Main Group (MG)	ಣ ಣಾ ಣಿ ಣೆ ಣೌ ಣು ಣೊ

Table 9. Examples of RA and BA stroke/stroke groups extracted from character combinations written together as one unit through segmentation, for training the recognition engine.

AG Group Characters	Stroke/Stroke groups retained from AG	
	RA	BA
ಸು	ು	–
ಸೂ	ೂ	–
ಯ	ಽ	–
ಯೌ	ೌ	–
ಣೇ	ೇ	–
ಣಂ	ಂ	–
ಣ್ಯ	್	ಯ
ಸ್ವ	್	ವ

$$F = [f_1, f_2, ... f_{30}]$$

of feature vectors

$$f_i = (a_i, b_i, r_i, \theta_i, a_{di}, b_{di})^T$$

Normalized horizontal and vertical coordinates: a_i and b_i

The size normalized horizontal and the vertical coordinates given by equations (3) and (4) are used as the first two features.

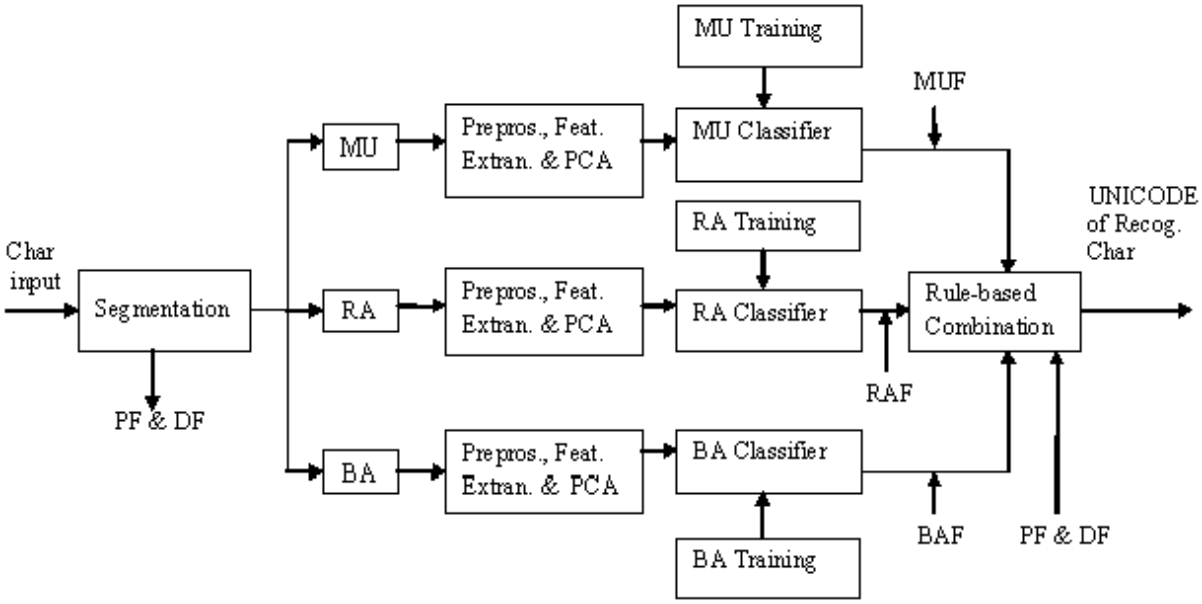

Figure 5. The proposed classification scheme

Normalized Trajectory Feature: r_i and θ_i

$$r_i = \frac{\sqrt{a_i^2 + b_i^2}}{\max(\sqrt{a_i^2 + b_i^2})} \quad (5)$$

$$\theta_i = \frac{\arg(a_i + jb_i)}{\max(\arg(a_i + jb_i))} \quad (6)$$

Normalized Deviation Features: a_{di} and b_{di}

$$a_{di} = (a_i - \mu_a)/\sigma_b \quad (7)$$

$$b_{di} = (b_i - \mu_b)/\sigma_a \quad (8)$$

where,

$$\mu_a = (1/30)\sum_{i=1}^{30} a_i \ , \quad \mu_b = (1/30)\sum_{i=1}^{30} b_i$$

$$\sigma_a = \sqrt{(1/29)\sum_{i=1}^{30}(\mu_a - a_i)^2} \quad \text{and}$$

$$\sigma_b = \sqrt{(1/29)\sum_{i=1}^{30}(\mu_b - b_i)^2}$$

7. PATTERN CLASSIFICATION SCHEME

The proposed online handwritten Kannada character recognition scheme has the training phase and the testing phase. Figure 5 illustrates the proposed recognition strategy that uses the divide and conquer technique. The collected data is divided into training and the test sets. For the simulation, around 11952 samples are used for training the recognition engine and around 5229 samples are subjected for testing.

7.1 Training

Online handwritten Kannada database includes the Main Group (MG) and the Auxiliary Group (AG). For training, the characters of all classes in MG are preprocessed, feature extracted and mapped to a sub-space of their respective classes using PCA [1]. As the intention of AG data collection is to extract RAs and BAs, they are passed through segmentation and the RA and the BA stroke groups are extracted leaving the rest of the strokes. These RA and BA stroke groups are preprocessed. The extracted features from these units are mapped to sub-space of their respective classes using PCA. The sub-space features of MA, RA and BA are trained separately.

7.2 Testing

Both MG and AG character groups are combined for testing. In addition to this, characters with more than one RA or BA are added to the test set, in order to validate the proposed segmentation scheme. This test data is passed through the segmentation process. The segmentation starts by checking the presence of the dot and the pulli strokes. If they are present, they are removed from the character and the corresponding flags are set. Next the presence of the RA and BA stroke groups are segmented and RA flag (RF) and BA flag (BF) are set accordingly. As there could be more than one RA and BA units,

their numbers are also updated with the status of the flags. Output of segmentation has three units, the MU, the RA and the BA.

The segmented units are preprocessed and the extracted features are mapped to sub-space using PCA. K-NN classifiers are used to classify these three units. The order of classification of stroke groups is MU, RA and BA. Depending on the number of RA and BA units the classification process is repeated in the same order. The status of the dot and the pulli flags are checked. The outputs of the classifiers are combined in the order of classification based on the rules of character combination. The recognized character is represented in Unicode.

8. EXPERIMENTS AND DISCUSSION

The performance of the proposed segmentation and recognition system for the online handwritten Kannada characters is studied for the writer independent case. The performance of the recognition systems with and without incorporating the divide and conquer strategy for two sets of features is given in Table 10. It is clear from the table that, the performance of the proposed recognition system is better than the recognition system without segmentation. To further validate the recognition capability of the proposed system, experimentation is carried out using a Tablet PC by asking the native writers to write the character combinations within the reduced set and the complex character combinations, which are not in the current reduced set. The performance of the system is fairly good in recognizing any Kannada character combinations.

Table 10 reveals that, in spite of only 282 class recognition problem, the maximum recognition accuracy attainable from the proposed system is around 81%. This may be due to the variation in writing styles across individuals, misclassification among stroke groups and not incorporating a robust classifier. Misclassification among Kannada characters is mainly due to similarity between the shapes of the stroke groups. A few pairs of characters which get misclassified are shown in Figure 5. The minor difference in shape could be with the variation of character trajectory either in horizontal or vertical direction. So, as part of our future research work, extraction of character specific shape features and designing of a robust classifier to boost the recognition accuracy of the proposed system will be addressed.

9. CONCLUSIONS

The paper presents a divide and conquer strategy for online handwritten Kannada character recognition. The structural and the spatio-temporal information of handwritten Kannada characters are exploited to segment a compound character into three distinct stroke units and recognize them separately to overcome the complexity of the huge number of character combinations. The performance of the proposed system is studied, with and without the divide and conquer technique.

10. ACKNOWLEDGMENTS

The authors thank Technology Development for Indian Languages (TDIL) Programme, Department of Information Technology, Government of India for funding the collection of the Kannada online handwritten data used in this work. Thanks are also due to Mr. Suresh Sundaram for suggestions in improving this paper.

11. REFERENCES

[1] Deepu V, Sriganesh M and Ramakrishnan A G. 2004. Principal Component Analysis for Online Handwritten Tamil Character Recognition. Int. Conf. Pattern Recog. 2 (2004), 327-330.

[2] K H Aparna, et al. 2004. Online Handwriting Recognition for Tamil. Proc. of the Int. Workshop on Frontiers in Handwriting Recognition. (2004), 438-443.

[3] Niranjan Joshi, G Sita, A G Ramakrishnan and M.Sriganesh. 2004. Comparison of Elastic Matching Algorithms for Online Tamil Handwritten Character Recognition. Proc. of the Int. Workshop on Frontiers in Handwriting Recognition. (2004), 444-449.

[4] Hariharan, Swethalakshmi, Anitha Jayaraman, et al. 2006. Online Handwritten Character Recognition of Devanagiri and Telugu Characters using Support Vector Machines. Proc. of the Int. Workshop on Frontiers in Handwriting Recognition. (2006).

[5] Anitha Jayaraman, C. Chandra Shekar, et al. 2007. Modular Approach to Recognition of Strokes in Telugu Script, Int. Conf. on Document Analysis and Recognition. (2007), 501-505.

[6] L Prashanth, V jagadeesh Babu, et al. 2007. Elastic Matching of Online Handwritten Tamil and Telugu Scripts using Local Features. Int. Conf. on Document Analysis and Recognition. (2007), 1028-1032.

[7] B VijayKumar and A G Ramakrishnan. 2004. Radial Basis Function and Subspace approach for Printed Kannada Test Recogntion. Int. Conf. Acoustics, Speech Signal Proc. (2004), 321—324.

[8] T V Ashwin, P S Sastry. A font and size-independent OCR system for printed Kannada documents using support vector machines, Sadhana, 27. (2002), 35-58.

[9] Ramakanth Kumar P. 2007. Off-line Handwritten Kannada Text Recognition using Support Machines using Zernike Moments, Intl. Conf. on Sys. Cybernetics Informatics. (2007), 626-632.

[10] S R Kunte and Sudhakar Samuel. 2000. Wavelet Features based Online Recognition of Handwritten Kannada Characters. Jl. Visualization Society of Japan, 20. (2000), 417-420.

[11] Ju-Wei Chentt et al. 1996. On-Line Handwriting Recognition of Chinese Characters via a Rule- Based Approach. Proc. of Int. Conf. on Pattern Recognition, 3. (1996), 220-224.

[12] Yong G E, et al. 2005. Online Chinese Character Recognition System with Handwritten Pinyin Input. Proc. of Int. Conf. on Document Analysis and Recognition. 2. (2005), 1265-1269.