ISSN:0975-9646

Handwritten Marathi Vowels Recognition using Correlation Coefficient

Parshuram M. Kambale
Department of Computer Science
Greenfingers College of Computer and Technology,
Yashwantnagar-Akluj, Maharashtra, India- 413118.

Balasaheb J. Kshirsagar
Department of Computer Science
Greenfingers College of Computer and Technology,
Yashwantnagar-Akluj, Maharashtra, India- 413118.

***Abstract*- In this paper we propose a handwritten Marathi vowels recognition technique based on correlation coefficient. Our algorithm accepts handwritten document containing Marathi characters as input taken from different individuals. This document is preprocessed to eliminate noise using morphological operations, and then line segmentation is performed to extract each line from the document. Each character from every line is segmented further. We use the correlation coefficient based template matching method to recognize the extracted characters. Correlation coefficient is computed between every extracted character and characters available in the database. The high value of correlation coefficient between any two characters indicate successful match of character. The proposed algorithm will yield encouraging results.**

***Keywords*— Marathi Vowels Recognition, Template Matching, Correlation Coefficient.**

I. INTRODUCTION

Marathi is one from the 23 official languages of India. It is an Indo-Aryan Language spoken by about 73 million people across the world. Marathi has some oldest journalism of all modern Indic, Indo-European language. The dialects of Marathi are called warhadi Marathi and standard Marathi. There are other sub-dialects which are Dangivadvali, Samavedi, Khandeshi, Ahirani, and Malwani. Marathi is the official language of State Maharashtra as well as Daman and Diu, Dadra and Nagar Haveli. Marathi language uses 60 phonemic letters, divided into three groups namely swear (Vowels 13 letters), Vyanjan (Consonants 34 letters) and Ank (numbers 10 digits) as well as Modifiers (Diacritics 12 letters). Development of offline and Online OCR for Marathi handwritten characters and numbers recognition is challenging work for researchers because while writing, characters have variation in shape, size, style of scripting, orientation, slant and different types of noise and strokes [10]. Few works are reported in literature for the recognition of Marathi and other Indian language characters and numbers. The character recognition task has been attempted by many researchers using techniques like template matching, statistical techniques and artificial neural network etc. Khanale and Chitnis [1] proposed the handwritten Devanagari characters recognition algorithm based on artificial neural network. They used a two layer feed forward neural network with 10 neurons and the log-sigmoid transfer function. This work reported 96% accuracy in recognition. Shelke and Apte [3] have proposed a multistage recognition technique for the recognition of handwritten Marathi compound character using multiple features. In their work firstly, they classified the compound characters using a two stage structural classification with different structural parameters. In the next stage, different features such as pixel density features, Euclidean distance and modified wavelet approximation features were obtained from characters which were structurally classified and normalized. These three features were fed to three different neural networks. The final recognition result was chosen based on majority voting. This work reported an accuracy of 97.95%. A work on structure based feature extraction of Marathi handwritten words is proposed by Mahender and kale [6]. They used rule based technique to recognize the words and claimed the accuracy of about 90%. The performance of the proposed method is poor when the size of vocabulary was increased. Tian Fu Gao and Cheng-Lin-Liu [7] Proposed Chinese Character recognition by using linear discriminant analysis based on compound distances. They used LDA for the estimation of discriminant vector for good discriminability and shown that under restrictive assumptions, the CMF acts as a special case of LDA based method. They evaluated the proposed methods by conducting experiments on the databases from ETL9B and CASIA by adapting the modified quadratic discriminant function (MQDF) treated as baseline classifier. Cheng-Lin-Liu, Masashi Kogo and Hiromichi Fujisawa [8] had recognized Japanese Handwritten character string by using the lexicon driven segmentation and lexicon matching. This work first introduces some kind of effective techniques for text line image pre-processing and pre-segmentation. Here the lexicon matching is used for consecutive segments to dynamically combine them into candidate character patterns. A classifier technique for characters is embedded in lexicon matching to obtain the characters which have matched with a candidate pattern from an available dynamic category set. This method reported correct rate of 83.68% and error rate less than 1%. A template matching based approach was proposed by Hegadi R. S. [2] for the recognition of Kannada numerals. This method uses the correlation coefficient for matching the numeral. The proposed method achieves an accuracy of 91%. In another work by Hegadi R. S. [5] a multilayer feed-forward neural network is used for the classification of printed Kannada numerals. The experimentation is carried out using all the existing fonts of printed Kannada numerals and this algorithm could recognize all the numerals. Pisal and Kamble [9] proposed the application of water fill feature extraction. In this paper we propose a coefficient correlation based feature extraction technique for recognition of Marathi handwritten vowels. Section 2 describes the proposed methodology, the feature extraction is discussed in section 3, results of the proposed method are discussed in

section 4 and conclusions are drawn in section 5. Patil, Adhiya and Ramteke [10] proposed the Moment Invariant Approach, Affine Moment Invariant Approach and combinations of both for Marathi vowel Recognition. The MI approach givse 75%, AMI gives 89.09% and combinations of MI and AMI gives 52.90% accuracy. Kakade and Raut [11] proposed performance analysis of handwritten devnagari characters using Machine Intelligence Approach. Here the performance of handwritten devnagari character is analyzed using different performance measures like Square error (MSE), percentage error (%E), performance characteristics and recognition percentage etc.

II. PROPOSED METHOD

Data collection for the experimental analysis has been taken from the different individuals for different age group. Database is created by scanning all the samples of handwritten vowels. Some sample scanned image of vowel is shown in below figure 1. Following are the steps applied during the recognition-

1. Image Acquisition
2. Image Pre-Processing
3. Segmentation
4. Template Matching
5. Feature Extraction
6. Output Image

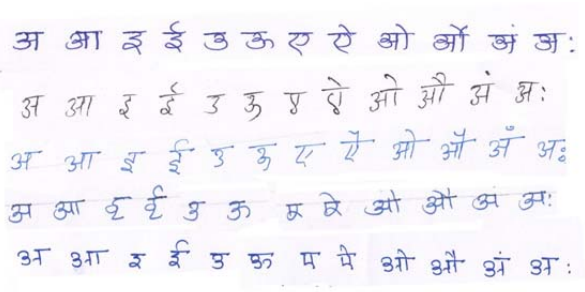

Fig. 1. Original scanned handwritten Marathi vowels

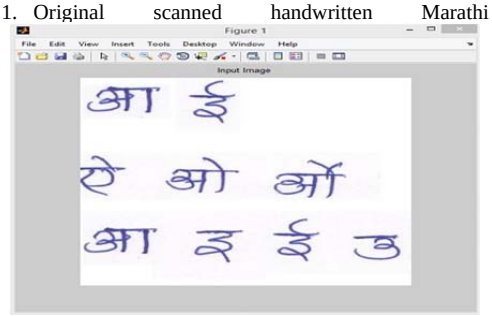

Fig. 2. Input Image

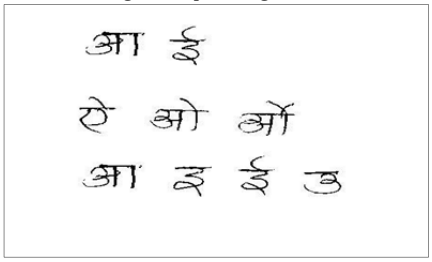

Fig. 3. . Binary image

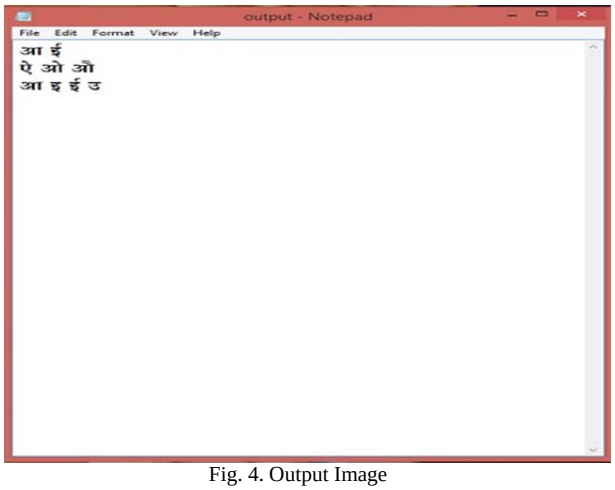

Fig. 4. Output Image

A. *Image Pre-Processing*

Marathi language has 12 vowels (Figure 1). The handwritten character image document containing Marathi characters is converted to binary image by applying thresholding shown in (Figure 3). The preprocessed image will contain the data in multiple lines, each lines containing multiple Characters. The document is scanned from top-left corner of the image and when pixel is found, its corresponding bottom-right point is located. The rectangle formed with the help of these points is extracted, which will be one line of text containing characters. The image document may contain many such lines. The extracted line is segmented into individual characters. Image by identifying the connected components and drawing the bounding box for each connected component. They are resized to standard size of 42 × 24 pixels for further processing.

B. *Template Matching*

The training data set for each character is collected by different 200 writers and stored in the database. The characters in this data set will have a size of 42 × 24 pixels. For each character a bounding box is drawn and the content within this box is stored. A correlation coefficient is computed between the segmented image and every character image data in the database. The correlation coefficient is a real value between -1 and 1. As the strength of the relation between the predicated values and actual values increases so does the correlation coefficient. A perfect fit gives a coefficient of 1.0.

III. FEATURE EXTRACTION

A training data set for each handwritten Marathi vowels characters is collected by different writers. The vowels in the dataset will have a size 42 × 24. For each character, a bounding box is drawn and the content within box is stored. A correlation coefficient is component between the segmented image and every characters image data in the database. The correlation coefficient is a real value between 1 and -1. As the strength of the relationship between the predicated values and actual values increases so does the correlation coefficient. A perfect fit gives a coefficient of 1. The mathematical formula for computing correlation coefficient value, p is.

$$p = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)2}\sqrt{n(\sum y^2) - (\sum y)2}}$$

Where n is the number of pairs of data, x and y are the training and testing character image data and p is correlation coefficient. The Correlation between the characters matches the template from the database. For each vowels character from the input image 12 correlation coefficient values are generated with each values corresponding to the characters. The maximum highest correlation coefficient values among these will show the matching characters from the database.

IV. Experimental Result

Experimentation is carried out on 10 different Marathi handwritten vowels sample documents. Figure 1 shows as input scanned image containing some handwritten Marathi vowels in rows. The segmented characters images are resized to 42 × 24 pixels. The correlation coefficient is computed between the training image data sets with the character. Testing character image data such results are recorded as shown in Table 1. In this table test sample represents total sample taken for recognition, correct classification represents number of samples recognize, and rate of recognition represents the percentage of result accuracy for each vowel. The Table 2 shows correlation of each handwritten Marathi vowels character.

TABLE I. Rate of Accuracy in Detecting Printed Marathi Vowels

Marathi Vowels	English Pronunciation	Test Sample	Correct classification	Rate of Recognition (%)
A	A	10	9	90
Aa	AA	10	10	100
[	E	10	8	80
[-	EE	10	7	70
]	U	10	9	90
}	OO	10	8	80
e	AE	10	8	80
eo	AU	10	7	70
Aao	O	10	7	70
AaO	AU	10	10	100
AM	AM	10	7	70
AÁ	Aha	10	7	70
	Average Accuracy			**80.83**

TABLE II. Correlation of Testing and Traning Data Set

	Correlation of Samples				
Marathi Vowels	**1**	**2**	**3**	**4**	**5**
A	0.9058	0.9058	0.60916	0.9058	0.9058
Aa	0.127	0.127	0.60578	0.127	0.127
[	0.9134	0.9134	0.6024	0.9134	0.9134
[-	0.6324	0.6324	0.59902	0.6324	0.6324
]	0.6362	0.5956	0.59564	0.6362	0.6362
}	0.64	0.5922	0.59226	0.64	0.64
e	0.62944	0.6294	0.58888	0.6294	0.6294
eo	0.62606	0.6260	0.9058	0.6260	0.6260
Aao	0.62268	0.6226	0.127	0.6226	0.6226
AaO	0.6193	0.6193	0.9134	0.6193	0.6193
AM	0.61592	0.6159	0.6324	0.6159	0.6159
AÁ	0.61254	0.6125	0.6362	0.6125	0.6125

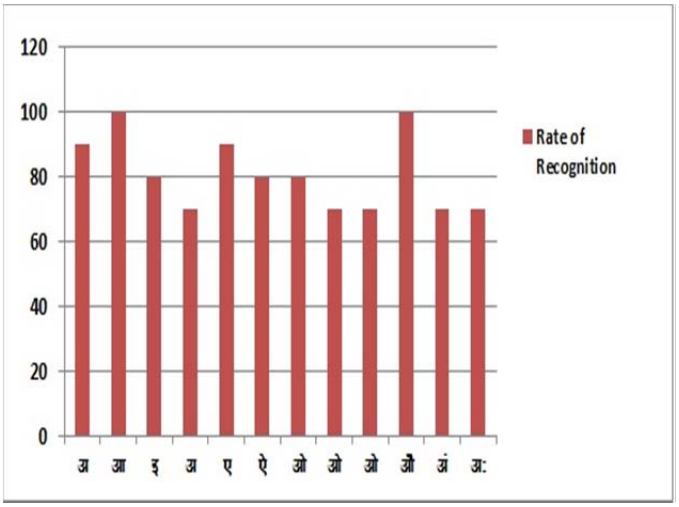

Fig. 5. Graphical Representation of Rate of Recognition of each Vowel

V. Conclusion

The template matching algorithms is old and well known, here an attempt is made to implement this algorithms based on the correlation coefficient to recognize and classify the handwritten Marathi vowels characters. The main characteristics of the Marathi characters are their shapes which are mostly formed with more curves. Most of the failures in recognition are due to either characters with sharp edges and corners, or breaking of characters making it as separate characters. In this paper, we have proposed a system for handwritten Marathi Vowel Recognition. The recognition accuracy by using correlation coefficient is promising, but more work is needed to be done. This handwritten Marathi vowel recognition (HMVR) system gives with almost 80.83% of recognition rate.

VI. FUTURE WORK

The above system has needed to improve over the following points-

1. Segmentation of characters including interconnected and overlapped.
2. Reduction in error rate.
3. Recognition of Consonants, numbers, and Diacritics.

REFERENCES

[1] R. S. Hegadi, "Template Matching Approch for Printed Kannada Numeral Recognition", Int. Conf. Comp. Int. Info. Tech, pp.480-483, 2011.

[2] R. S. Hegadi, “Classification of Kannada Numerals Using Multi-layer Neural Network”, Adv. in Int. Sy. and Comp., Vol. 174, pp. 963-968, 2012.

[3] T. B. Pisal and P.M. Kamble "Marathi Character Recognition by using Probablistc Neural netwok Classification", Int. Jr. Comp. Sci. Info. Tech, pp.66-63, 2012.

[4] P. B. Khanale and S. D. chitins, “Handwritten Devanagari Character Recognition Using Artificial Neural Network”, Journal of Artificial Intelligence, vol. 4, no. 1, pp. 55-62, 2011.

[5] S. Shelke and S. Apte, “Multistage Handwritten Marathi Compound Character Recognition Using Neural Networks”, Journal of Pattern Recognition Research., vol. 2, pp. 253-268, 2011.

[6] P. E. Ajmire, R. V. Dharaskar and V. M. Thakare, “A Comparative Study of Handwritten Marathi Character Recognition”, Int. J. of Computer Applications, pp. 26-28, 2012.

[7] C. N. Mahender and K. V. Kale, “Structured based Feature Extraction of Handwritten Marathi Word”, Int. J. of Comp. Applications., Vol. 16, no. 6, pp. 42-47, 2011.

[8] Tian-Fu Gao and Cheng-Lin Liu, “High accuracy handwritten Chinese character Recognition using LDA-based compound distances”, Pattern Recognition, vol. 41, no. 11, pp. 3442-3451, 2008.

[9] Cheng-Lin Liu, “Lexicon-Driven Segmentation and Recognition of Hand-written Character Strings for Japanese Address Reading”, IEEE.Tran. On Pattern Analysis and Machine Intelligence., Vol. 24, no. 11, pp. 1427-1437, 2002.

[10] N. P. Patil, K.P. Adhiya and S.P. Ramteke, “A Strucutred Approach to Handwritten Marathi Vovel Recognition”, International Journal of Computer Applications, vol. 31-No. 3, pp. 48-52, 2011.

[11] P. M. Kakde, V. R. Raut, “Performance and Analysis of Handwritten Devnagari Characters Recognition Through Machine Intelligence”, International Journal of Emerging Techology and Advanced Enginnering, Vol. 2, Issue 7, pp. 241-247, July 2012.