Database Development and Preprocessing for Handwritten Marathi Numeral Recognition

Mr.Rohit R. Pawar.
M.Sc. Student
Dept Of Computer science,
Solapur University Solapur

Ms.Minakshi S.Bhandare
M.sc.Student
Dept.Of Computer Science
Solapur University Solapur

Ms.Chaitali D.Koulage.
M.sc.Student
Dept.Of Computer Science
Solapur University Solapur

ABSTRACT

This paper applies different techniques for recognizing printed and handwritten Marathi numerals. The concentration is done on work carried out on different data sets which is collected by different peoples. We studied most of the published papers related to this topic and from this we analyze different methodologies and their results. It gives proper guidance for research in OCR. In this paper we work dealing with recognition of numerals. Our research work is for handwritten numerals recognition from scanned documents. We know that by selecting proper technique high performance and better accuracy will get in recognition. The aim of this paper to improve recognition result by applying different types of techniques on data sets .It also focuses on problems related to the recognition of numerals and also locates the researchers the way where there is still a scope of accuracy.

General Terms

 Pattern Recognition.

Keywords

Marathi Numerals Recognition, OCR.

1. INTRODUCTION

In this paper we discusses about handwritten Marathi numerals. Recognition of Marathi numerals is one of the biggest problems in present scenario. Recognition of handwritten numerals is challenging and most difficult task in pattern recognition because of its complex and structural properties. The reason behind this complex structure is that different writers have different writing styles. Because of these variations in writing styles the job becomes difficult.

1.1 Types of Handwriting Recognition:-

Handwritten recognition technique is basically divided in to two types.

1.1.1 Online handwriting recognition:-

In online recognition system, the system recognize continuously as the writers writes the data. It basically goes along writing process.

1.1.2 Offline handwriting recognition:-

In offline recognition system operations are performed after the writing or printing is complete [1].

1.2 Problems in handwriting recognition:-

 Sometimes numerals are wrongly classified because of many reasons listed below.

 Numerals are not properly written.

 Some numerals have similar structures.

Sometimes two or more numerals may overlap.

Sometimes system confuses the numerals with some other numerals and does not recognize it correctly. [1]

Fig 1: Some of printed Marathi Numerals

2. LITERATURE SURVEY

The work on Handwritten Devanagari numerals is carried by Mr. Shailendra Kumar Shrivastava and Mr. Sanjay S Gharde [2] proposed Support Vector Machine for Handwritten Devanagari Numerals Recognition in which Support Vector Machine (SVM),Moment Invariant ,Affine Moment Invariant techniqes are used.Mr Akhilesh Pandey,Amresh Kumar,Rajiv Kumar and Amod Tiwari [3] introduced Handwritten Devanagari Number Recognition using Majority Voting Scheme apply Majority Voting Scheme for off-line hand written numbers. Mr. Shashi Kant Shukla and Akhilesh Pandey [4] have developed Classification of Devanagari Numerals using Multiple Classifier for classification strategies to recognize unconstrained handwritten numbers.The work on Kannada Numerals and its classification using multi-layer neural network proposed by R.S.Hegadi [5]. First system for hand-written numeral recognition of Devanagari characters was proposed by R.Bajaj [6]. A Template matching approach on Kannada numerals recognition done by R. S. Hegadi [7]. R.S.Hegadi and P.M.Kamble [8] introduced the technique on Marathi Handwritten Numerals recognition Using Multi-Layer Feed-Forward Neural Network. P. M. Patil and T.R. Sontakke [9] also presented an algorithm for handwritten Devanagari numeral recognition which was based on rotation, scale invariant

3. ABOUT DATABASE

Research in handwritten Marathi numerals recognition divided in to two categories .i.e. Online and Offline. In online systems camera is used. Camera captures the images with help of sensor while writing process is executing and then makes that information dynamically available according to the strokes. Offline systems work statically. In this images are scanned after completion of writing processes on paper sheets.
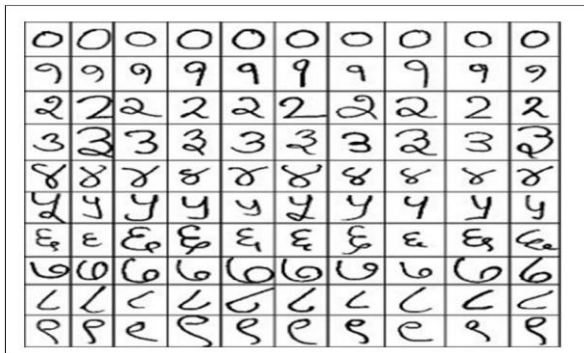
Fig 2: Existing data of handwritten Marathi numerals.

4. EXISTING SYSTEM

The existing system contains the following results which are done by some researchers. Feature extractor techniques used to extract features from each image i.e. up to 18 features that are used in SVM for recognition purpose. Implementation of Binary classification techniques of SVM is done and linear kernel function is used. This linear Support Vector Machine produces 99.48% recognition rate [2]. The feature extraction and majority voting scheme technique are uses simple profile and contour based triangular representation of technique for back propagation and cascade feed forward neural network for classification. The average recognition result of this approach is 94.16% [3]. Using multiple classifier schemes over KNN new neural network, LDA, and KNN techniques are used for feature extraction and classification. The KNN classifier will produce the result 93.82% and LDA classifier will produce 92.97% [4].
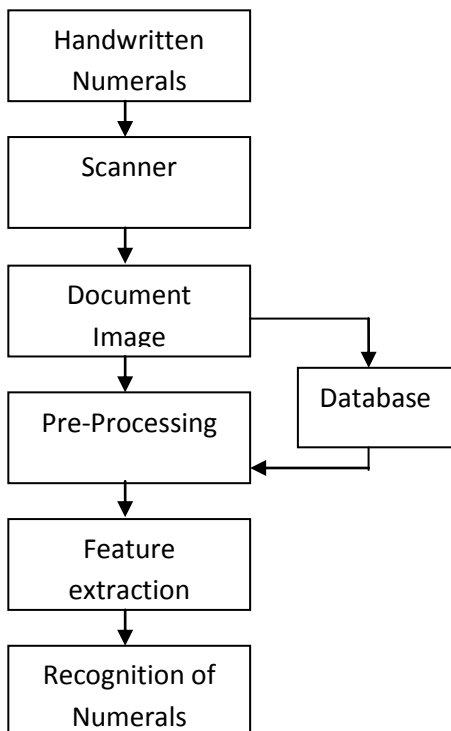
5. PROPOSED SYSTEM

Fig 3: Block Diagram of Numeral Recognition System

5.1 Database Collection

In this paper we use the OCR which is an offline process which helps us for analyzing static documents. Here we are taken 10 digits and each paper sheet contains 250 sample digits which are used for work. These data sets are written by different age group peoples. After getting these samples we scanned these documents and save it in jpg file format.
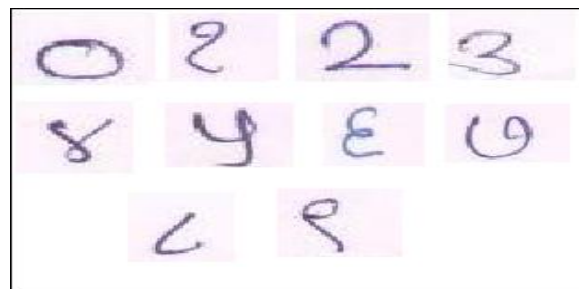

Fig 4: Collected Numeral database

Fig 5: Some cropped images

Here we are taken 250 samples for each digit is cropped with the help of program which is written in MATLAB.

5.2 Pre-processing

Pre-processing is most important step. Here first we store collected images in to database and perform following steps on each image.

5.2.1 Taking input image for operation

Fig 6: Original Image

5.2.2 Gray image

Fig 7: Gray Image

5.2.3 Thresholding

Next step is thresholding in which all pixels converted in to white above the threshold and black below side. It is also binarization process.

Fig 8: Thresholding Image

5.2.4 To remove the noise from image filtering is used.

Fig 9: Filtering of image

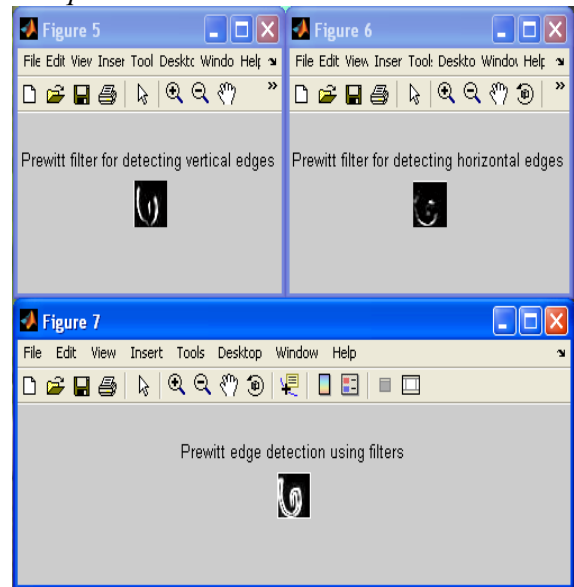### 5.2.5 Edges are detected using edge detection technique

Fig 10: Edge detection

5.2.6 Bounding Box

For extracting features from these numerals we use boundary descriptor i.e. bounding box .using which we get all features of the images.
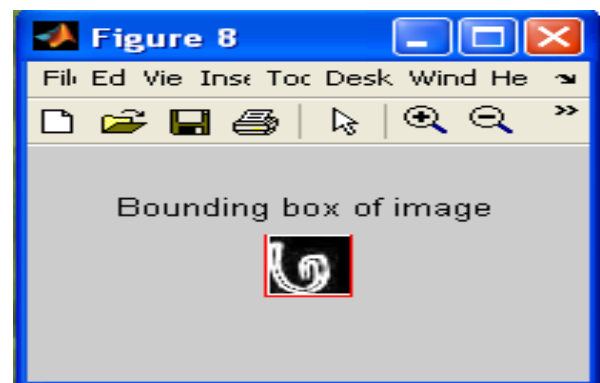

Fig 11: Bounding box of Image

Following are the features which are calculated with the help of bounding box.

Area- It calculates the region of pixels in numbers.

Major Axis Length - It calculates the length of the major axis of the ellipse.

Minor Axis Length- It calculates the length of the minor axis of the ellipse.

Eccentricity- It calculates the interval between the foci of the ellipse also compute its major axis length and ratio.

Euler Number-It finds difference between number of object in the region and number of holes in the objects.

Extrema- It consist extrema points in the region.

Solidity- It is computed using the formula Area/Convex Area

Equiv Diameter- It is computed using the formula

sqrt (4*Area/pi).

Area: 848

Centroid: [14.2123 18.0366]

Bounding Box: [0.5000 0.5000 30 33]

Sub array Id x: {[1x33 double] [1x30 double]}

Major Axis Length: 37.9384

Minor Axis Length: 32.8837

Eccentricity: 0.4987

Orientation: -76.6367

Convex Hull: [9x2 double]

Convex Image: [33x30 logical]

Convex Area: 990

Image: [33x30 logical]

Filled Image: [33x30 logical]

Filled Area: 990

Euler Number: -5

Extrema: [8x2 double]

Equiv Diameter: 32.8589

Solidity: 0.8566

Extent: 0.8566 Pixel Id x List: [848x1 double]

Pixel List: [848x2 double]

6. RESULT AND DISCUSSION

This paper has shown the result of pre-processing and feature extraction of Marathi numerals. Here we collected up to 250 samples of per number. Data is collected from various age groups of writers. The initial procedure of pre-processing and feature extraction is done. This is used for classification and recognition of numerals.

7. CONCLUSION

From our work we conclude that the techniques used for the pre-processing as well as feature extraction of numeral characters are easier. We applied those techniques on our database which gives us better result. In future we are going to apply various techniques like Template matching, Vector feature, chain code approach on reorganization of numerals. The recognition result will provides new benchmark for recent and future research.

8. ACKNOWLEDGEMENTS

9. REFERENCES

[1] Madhav Goyal, Naresh Kumar Garg,"Survey Paper on Handwritten Devanagari Numeral Recognition",IJAIR vol.2 Issue 6 ISSN:2278-7844

[2] Shailendra Kumar shrivastva ,Sanjay S. Gharde,"Suport vector machine for Handwritten Devanagari Numerals Recognition" ,IJCA(0975-8887),vol.7-No.11,Oct 2010.

[3] Akhilesh Pandey,Amresh Kumar,Rajiv Kumar and Amod Tiwari,"Handwritten Devanagari Number Recognition using Majority voting scheme",IRACST-International journal of computer science and Information Technology and Security(IJCSITS),ISSN :2249-9555 vol.2, No.3,June 2012.

[4] Shashi Kant Shukla, Akhilesh pandey,"Classification of Devanagari numerals using Multiple classifier", International Journal of Computer Trends and Ttechnology(IJCTT)-vol.12, No.4, June 2014.

[5] R.S.Hegadi "Classification of Kannada Numerals using Multi-Layer Neural Network",Adv. In Int Sy and Conf Comp Vol.174, PP.963-968, 2012.

[6] R.Bajaj, L.Dey, S.Chaudhary,"Devnagari Numeral Recognition by Combining Decision of Multiple Connectionist Classifiers", 2002.

[7] R.S.Hegadi,"Template Matching Approach for Printed KannadaNumeralsRecognition",Int.Conf.Comp.Int.Info. Tech,PP.480-483,2011

[8] R.S.Hegadi and P.M.Kamble,"Recognition of Marathi Handwritten Numerals Using Multi-Layer Feed-Forward Neural Network"IEEE Explore,WCCT PP.21-24,2014.

[9] P.M.Patil, T.R.Sontakke,"Rotation, Scale and Translation Invariant Handwritten Devnagari Numeral Character Recognition using General Fuzzy Neural Network", Pattern Recognition, vol.40, 2007.