International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Feature Extraction for Recognizing MODI Characters

Anjali S. Bhalake[1], R. S. Hegadi[2]

[1]Dept. of Comp. Science, Solapur University, Solapur
anjalibhalake@gmail.com

[2]Dept. of Comp. Science, Solapur University, Solapur
rshegadi@gmail.com

Abstract: ***In this paper, Feature Extraction for recognizing modi characters are done, for this paper I referred some research paper and studied them. Then I compared their methodologies and technique used to recognize modi characters. For extracting features from modi characters I used Boundary descriptor.***

Keywords: Character Recognition, Boundary descriptor**,** Bounding Box.

1. Introduction

The modi word derived from the Marathi verb modane, which means to break. In 19th century modi script was popular for writing Marathi but modi script was very difficult for writing because of its cursive type so that devnagari script were developed. Using modi script reading historical documents and papers are very difficult and complicated. The Modi script already existed in the 1200s. It was introduced as an official script for Marathi by Hemadpant. He brought modi script from shrilanka. The Modi script was frequently used as a shorthand script. It was used primarily by administrative people in keeping their accounts and writing credit notes. It was also used to encrypt the message.

There are various styles of modi characters associated in particular era and many changes occurred in each era.

Proto Modi - The Proto Modi style used until 12th century.
Yadav Era – The Yadav Era style was used until 13th century.
Bahamani Era – Bahamani Era style was used until 14-15th century.
Shiva Era – Shiva Era style was used until 17th century.
Peshwa Era – Peshwa Era Style was used until 18th century.
English Era – It was the last era when Modi script was used.

Figure 1: Modi Consonants

Figure 2: Modi Vowels

Figure 3 : Modi Numerals

2. Existing System

As modi is cursive type, it is difficult to recognize modi characters. There are very few researchers who worked on modi characters. Following are the methods used to recognize modi characters.

2.1 Chain Code Approach

D.N.Besekar used Chain code and two layer feed forward network to recognize modi vowels. For Data preprocessing they have used median filter which reduce salt and paper noise of segmented character images. Then the character images are normalized to 56x56 using bicubic interpolation. Then for feature extraction they have calculated chain code. They got 65% to 75% accuracy to recognize modi vowels. They conclude that centroid of the image is used as additional feature which improve the result.

2.2 Hu, Zernike Moments and Zoning

Sadanand A. Kulkarni, Prashant L. Borde, Ramesh R. Manza, Pravin L. Yannawar used Hu, Zernike moment and zoning to recognize modi characters. For data preprocessing they performed morphological opening and closing and for extracting small elements Top Hat

Transform is used. Then for feature extraction they have used Hu's seven moments, Zernike moments and Zernike moments with zoning. They got 70-80% of accuracy in recognizing modi characters and they conclude that the accuracy with Hu's moment is 71.52, with Zernike moment is 76.74% and with zoning it is up to 82.61.

2.3 Comparison Table

Table1: Comparison Table

Researchers	*Method*	*Result*
D.N.Besekar	Chain code and two layer feed forward network (for vowels)	Accuracy is near about 65%-75%
SadanandA.Kulkarni Prashant L.Borde R.Manza, Pravin L.Yannawar	Hu,Zernike moments with zoning	Accuracy is near about 70%-80%

3. Proposed System

In proposed system for extracting features from modi characters I used boundary descriptor bounding box. By using this bounding box I calculated all its scalar points like area, majoraxislenth, minoraxislenth, etc.

3.1 Preprocessing

First of all for database collection I collected images of modi characters then segmented each character and stored them in database. Then by taking one by one character applied following steps

1. Take input image

Figure 4: Input image

2. Converted to binary image

Figure 5: Binary image

3. Noise is removed using median filter

Figure 6: After removing noise

4. Edges are detected using canny edge detection

Figure 7: Edge Detection

3.2 Feature Extraction

For extracting features from these characters I used boundary descriptor i.e. Bounding Box as shown in below.

Figure 8 : Bounding Box

Following are the features calculated from bounding box

```
            Area: 774
        Centroid: [38.8475 30.4031]
     BoundingBox: [9.5000 8.5000 56 47]
     SubarrayIdx: ([1x47 double]  [1x56 double])
 MajorAxisLength: 57.5865
 MinorAxisLength: 48.3543
    Eccentricity: 0.5431
     Orientation: -31.1183
      ConvexHull: [19x2 double]
     ConvexImage: [47x56 logical]
      ConvexArea: 1830
           Image: [47x56 logical]
     FilledImage: [47x56 logical]
      FilledArea: 1613
     EulerNumber: -38
         Extrema: [8x2 double]
   EquivDiameter: 31.3925
        Solidity: 0.4230
          Extent: 0.2941
    PixelIdxList: [774x1 double]
       PixelList: [774x2 double]
```

Figure 9: Feature Extracted Points

Area → It calculates the actual number of pixels in the region.
MajorAxisLenth→It calculates the length of the major axis of the ellipse.
MinorAxisLength→It calculates the length of the minor axis of the ellipse.
Eccentricity→ It calculates the ratio of the distance between the foci of the ellipse and its major axis length.
EulerNumber→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/ConvexArea
EquivDiameter→It is computed using the formula sqrt (4*Area/pi)

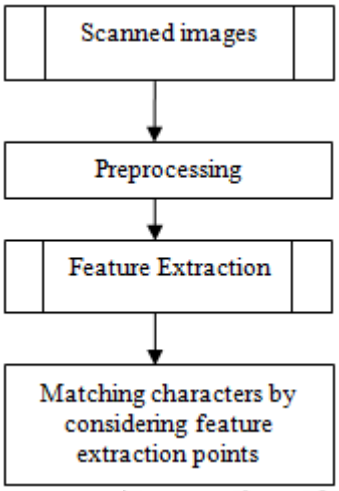

Figure 10: Block Diagram of Feature Extraction

4. Conclusion

By using these features we can recognize modi characters Millions of modi documents were waiting to unfold the history of Maratha History. There are very few people who know the modi script. There is need to work on modi characters so we can get more knowledge about our history.

5. Acknowledgements

I would like to thanks our guide Dr. R. S. Hegadi for supporting and encouraging for doing this research work. Their guidance is very helpful to me to the research work.

References

[1] http://Modi Lipi or Modi Script History of Modi Lipi.html
[2] Modi alphabet-Wikipedia, the free encyclopedia.html
[3] "Recognition of numerals of modi script using morphological approach"
[4] D. N. Besekar International Referred Research Journal, April, 2011.
[5] Global Online Electronic International Interdisciplinary Research Journal (GOEIIRJ) June 2013.
[6] Proposal to Encode the Modi Script in ISO/IEC 10646 by Anshuman Pandey November 5, 2011
[7] Offline Handwritten MODI Character Recognition Using HU, Zernike Moments and Zoning Sadanand A. Kulkarni1, Prashant L. Borde2, Ramesh R. Manza3, Pravin L. Yannawar4 1.2.4 Vision and Intelligent System Lab
[8] A Chain Code Approach for Recognizing Modi Script Numerals by D.N.Besekar and Dr. R.J. Ramteke
[9] Special Approach for Recognition of Handwritten MODI Script's Vowels by D. N. Besekar Proceedings published by International Journal of Computer Applications® (IJCA)
[10] http://Modi documents of Maratha era to be digitized – TheHindu.html