International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 08, 2020
ISSN: 1475-7192

Online Handwritten Telugu Stroke Recognition using Direction Operator for Feature Extraction

[1]Srilakshmi Inuganti, [2]Dr. R. Rajeshwara Rao

Abstract--*Important phase of Character Recognition is Feature extraction. In this paper a feature extraction technique using 8-point interconnection method named as Direction operator is proposed. The feature obtained by using direction operator represents structural, global and local characteristics of a stroke. This feature represents (x,y) coordinate of each point on the stroke with 8-bit vector, where each vector gives interconnection of this point with other points in the stroke in 8-directions.A two stage classifier known as Hybrid classifier is also proposed, in which direction operator feature is used in preclassifier and preprocessed(x,y) coordinates are used in postclassifier. We have verified this feature with HP-Lab data available in UNIPEN format. Experimental results proved that direction operator feature improves recognition accuracy over the chosen dataset.*

Key Words--*Direction operator, Hybrid Approach, Stroke Recognition, Online*

I. INTRODUCTION

With the rapid advancements in the handheld devices like PDAs, mobile phones etc., there is an increasing interest in Online Handwritten Character Recognition (OHCR) research. OHCR would be the better choice for these tiny devices when compared to traditional keyboards.It is the process of recognizing a stroke as we are writing on the device with a pen.Stroke is an instance between pen down and pen up on the device. The steps involved in OHCR are: preprocessing, feature extraction and recognition. Feature extraction plays vital role in character recognition, especially in case of Telugu language, which has huge character setand many of the characters resemble one another in structure [1]. Extracted featuresare given as input to classifier, which affects the recognition accuracy significantly. Any online handwritten character can be described by local or global or structural features or combinations of these.

Local Features are those which are extracted at each point on the stroke. A stroke is a direction of pen down to pen up, so these features represent the relationship between consequent points of the trajectory.The basic local feature is xy-coordinate of each point in the trajectory[2]. Assamese Online Handwritten digit is recognized using features (x,y), (Δx,Δy), (ΔΔx,ΔΔy) and distance between end points is de.In this temporal change (Δ) and how fast the temporal change itself is happening (ΔΔ) among consequent (x,y) coordinates are represented and they are unique for each numeral[3].Another popular feature is tangent slope angle feature, which gives directional data instead of simple linear data [4]. Direction element feature [5] is constructed from dot-orientation of pixels in character. Here to find the direction of each black pixel, eight surrounding pixels are used. This method unable to utilize dynamic information of a character. Trajectory and velocity modelling are proposed to represent handwriting system. These are sequence of overlapped Beta functions. The parameters proposed here are used as local

[1]*Research Scholar, Computer Science and Engineering, Jawaharlal Nehru technological university, inuganti.srilakshmi@gmail.com*
[2]*Professor, Computer Science and Engineering, Jawaharlal Nehru technological university, University college of engineering, vijayanagaram, raob4u@yahoo.com*

DOI: 10.37200/IJPR/V24I8/PR280053
Received: 19 Jan 2020 | Revised: 06 Feb 2020 | Accepted: 12 Mar 2020

features[6]. However local features are not strong enough to overcome shape distortion,specifically fluctuation in stroke direction,rotation, and slanting usually found in handwritten characters.

Global features give the coordination between two arbitrary points on the stroke.These have the capability to represent numerous important characteristics of a stroke. Importance of Global features described and confirmed in [7]. Delaunay triangular discriminator [8], which is a good context discriminator, used as a feature vector for a handwritten stroke. It covers most global features, so increases discrimination power, but constructing Delaunay triangles takes more time. Structural and shape of a Kanji stroke are provided by using 2D arrangement of adjacent strokes[9].The collection of 18 shapes are usedto represent Tamil Characters [10],these are not applicable to distortion occurred in hand writing.Global features of handwritten Tamil characters are represented using Discrete cosine transform and Discrete Fourier transform. Usually confusing pairs are distinguishable by a small region. To overcome this problem discriminative region identification method proposed in[11] .This identifies the discriminative region that discriminates between confusing pairs. Here 14 different features are used. Complete information about the Stroke are represented using Global descriptors. However, they often unable to work well with similar classes and confusing pairs that have minor point-wise variations.On the other side, local descriptors are extracted at every coordinate and so these give desirable inter-class separation.Critical points of a stroke are identified and angular variation between these points is calculated in[12],which captures local and global features of a stroke.Here finding appropriate number of critical points is a tedious task.To utlize the best of both features, we experimented with the concatenation of local and global features. In this paper we propose feature using direction operator which captures local, global and structural features of stroke. We compare the performance of proposed feature with simple (x,y) coordinate feature and (x,y) coordinate & tangent feature over theHP-Lab data available in UNIPEN format. The rest of the paper is as follows. Details of proposed approach are illustrated in Section 2. Overview of recognition process depicted in Section 3.Experimental results and comparisons are reported in Section 4.In Section 5, our findings are summarized.

II.A NEW FEATURE: DIRECTION OPERATOR FEATURE

The Direction operator is applied at each (x,y) coordinate to obtain 8-bit vector corresponds to the (x,y) coordinate. The 8-bit vector gives interconnection points in the stroke in 8-directions with respect to selected (x,y) coordinate[13]. The direction operator is as represented in Figure 1. At every coordinate in the stroke this operator is used. It representschosen point with respect to remaining coordinates in the trajectory in eight directions.

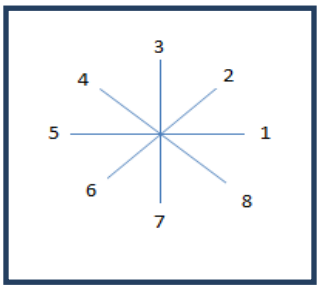

Figure 1: Direction Operator

DOI: 10.37200/IJPR/V24I8/PR280053
Received: 19 Jan 2020 | Revised: 06 Feb 2020 | Accepted: 12 Mar 2020

2.1 Definition for Direction operator:

For each coordinate X_i on a stroke, the direction operatorgiven in Figure 1 isused to find out the intersections in 8-directions. An 8-bit vector is obtained for each point on the stroke, where each bit is set to 1 if there is an interconnection point in that direction. Figure 2 illustrates application of direction operator for a stroke at point X_i. An 8-bit vector for every point X_i is calculates as follows:

1. Identify the points of intersection to each line of direction operator in all the quadrants after using the direction operator at the point X_i.
2. If the identified nearest points in quadrant 1 and 4 for line 1 are consequent points, then first bit in 8-bit vector is set to 1.
3. If the identified nearest points in quadrant 1 for line 2 are consequent points, then second bit in 8-bit vector is set to 1.
4. If the identified nearest points in quadrant 2 and 3 for line 5 are consequent points, then fifth bit in 8-bit vector is set to 1.

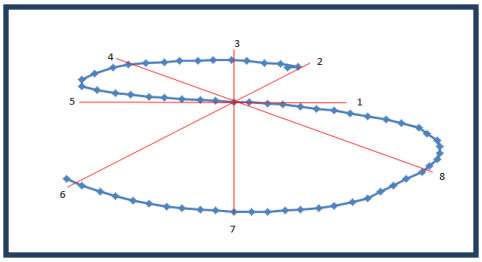

Figure 2: Applying Direction Operator

2.2 Algorithm for Feature Extraction

1. Apply the direction operator at each coordinate on the stroke.
2. Find the points of intersection in 8-direction lines with respect to selected point.
3. Fora feature vector FV_i is of size 8 bit,each bit is set to 1 if the nearest points find in that direction are consequent points.
4. To form a feature vector FV, concatenate FV_i obtained for all coordinates on the stroke.

III. OVERVIEW OF RECOGNITION PROCESS

3.1 Database

In a research area related to pattern recognition Benchmarking database is very important. In Telugu the data set available is Hp-Labs data in UNIPEN format. HP-Labs UNIPEN dataset contains 166 Telugu Characters each of 270 samples collected from 146 users in two trails[1].In the data collection process AcecadDigimemo device is used with Digimemo-DCT application. Among these collected 45,219 samples, 6640 samplescollected from 20 users is used for training and 3320 collected from 10 users used for testing.

DOI: 10.37200/IJPR/V24I8/PR280053
Received: 19 Jan 2020 | Revised: 06 Feb 2020 | Accepted: 12 Mar 2020

3.2 Preprocessing

Because of variants in size of writing strokes, it is necessary to transform a stroke into a standard window. In our database we have normalized to window size 200X250. Speed of handwriting results in missing points, these missing points can be calculated using linear interpolation. Later resampling is applied to obtain constant number of points, which are uniformly spaced. All samples in our database contains 65 points after Uniform resampling[15].Sample preprocessed character is shown in figure 3.

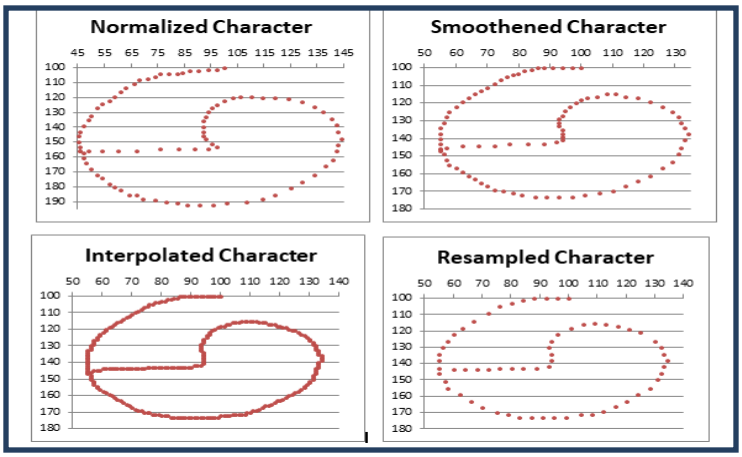

Figure 3: Preproccesed Character

3.3 Feature Extraction

The three features used are:

Pen-tip position: The preprocessed (x,y) coordinates are taken as pen-tip positions.

P= $\{p_i\}$ where i=0,1,…..,64

and $p_i = (x_i, y_i)$

Tangent Angle: This feature gives direction at pen-tip position. The tangent angle is calculated between two adjacent points of P as

$\theta_i = \tan^{-1}(\frac{y_{i+1}-y_i}{x_{i+1}-x_i})$where i=0,1,……64

The tangent angle at $\theta_{64} = \theta_{63}$

Direction operator Feature: The features extracted from the character by applying direction operator at each point on the stroke as described in Section 2. The vectors obtained for points P_1,P_2,P_3 and P_4 shown in Figure 4,are F_1=[1,1,1,1,1,1,1,1],F_2=[0,0,0,0,0,1,1,1],F_3=[1,1,1,0,0,0,0,0] and F_4=[1,1,0,0,0,0,1,1].All bits in F_1 are set to 1 ,as P_1 is inside a loop. Thelastthreebits in the vector F_2 are 1's , as there are intersection points in these directions only. FV = [F_1, F_2,...,F_N] where, N is the number of points in the character, ie feature vector is the concatenation of binary vectors obtained at every point on the trajectory.In our method each stroke results in 65 points after uniform normalization. Hence the size of our feature vector is 520(65 x 8) bits. Here the feature is extracted by finding the nearest neighbors of each point in each of the eight directions. This feature captures not only relative points to each

DOI: 10.37200/IJPR/V24I8/PR280053
Received: 19 Jan 2020 | Revised: 06 Feb 2020 | Accepted: 12 Mar 2020

point, but also direction information in 8-directions.This proposed feature can also be used to illustrate structure of the stroke.

3.4 Classification and Evaluation

In this section, experimental results with the above mentioned features are presented. Here four different classification approaches are used. All approaches are based on DTW,which allowscomparing two vectors of variable lengths.DTW is particularly helpful to compare vectors, where rate of progression is not linear.

3.4.1 Basic Approaches

The following approaches are basic as they are using single phase for recognition.

Approach 1

This approach uses simple (x,y) co-ordinates as features and nearest neighbor with DTW distance measure used as recognizer.

Approach 2

This scheme includes tangent angle between adjacent points in addition to simple (x,y) coordinates as features. Nearest neighbor with DTW distance measure used as recognizer.

Approach 3

This approach extracts features by applying direction operator at each point in the stroke. Nearest neighbor with DTW distance measure used as recognizer.

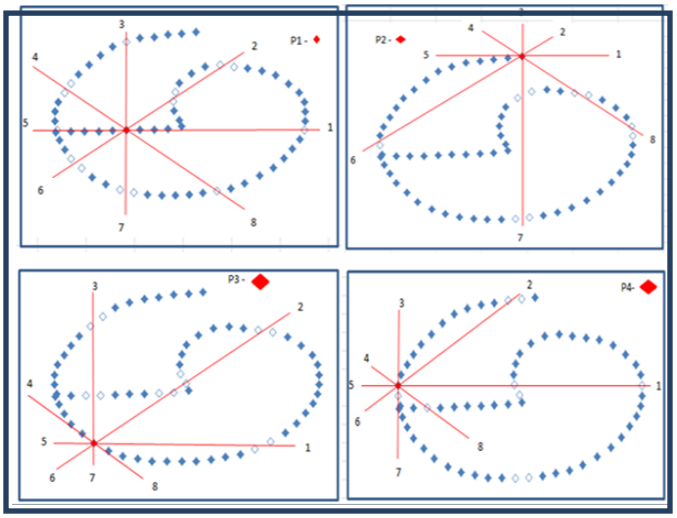

Figure 4: Direction operator at different point for character 'ഌ'

DOI: 10.37200/IJPR/V24I8/PR280053
Received: 19 Jan 2020 | Revised: 06 Feb 2020 | Accepted: 12 Mar 2020

3.4.2 Hybrid Approach

The following hybrid approach achieves the recognition task in two phases. In the primary phase pre-classification is performed and obtains top 10 choices as its output. Secondary phaseobtains the output from these 10 choices and provides post-classification.

Approach 4

In this approachPre classification utilizes features extracted by applying direction operator at each coordinate. Next preprocessed (x,y) coordinates used as features at post classification stage. The classification method used at both stages is 10-nearest neighbor and 1-nearest neighbor respectively.

IV. EXPERIMENTAL RESULTS

All experiments are conducted on the Telugu Database of HP Labs collected from 146 users for 166 characters. Among these collected 45,219 samples, 6640 samples collected from 20 users isused for training and 3320 collected from 10 users used for testing. The parameter evaluated for each feature is recognition accuracy and speed. Our main intention is to find performance of direction operator. Here the experiments are carried out as follows: first the strokes are preprocessed. Then, over the preprocessed stroke three different feature extraction methods are applied. We observe that recognition accuracy for approach 3 is more as the new features captures structural, global and local characteristics of stroke.However, recognition speed is less than remaining two approaches. This is due to the length of feature obtained in approach3 .The table 1 shows the results of nearest neighbor classifier using three basic schemes with different training sample sizes. Next our proposed hybridapproach is evaluated. Figures 5 and 6 depict the results of hybrid approach in terms of speed and accuracy, respectively, against different sample sizes. Hybrid approach increases the recognition speed, as the direction operator feature used in preclassifier contains only binary values. In addition to increased recognition speed, hybrid approach achieves recognition accuracy in par with approach 3.

Table 1: Percentage of Accuracy for basic approaches (6640 samplesfor training, 3320samples for testing)

Sample Size	**Scheme**		
	1	**2**	**3**
1500	66	68	81
2500	68	69	81
3500	69	70	83
4500	70	71	84
5500	71	73	84
6500	73	73	87
Speed	0.8 ch/sec	0.6 ch/sec	0.45 ch/sec

DOI: 10.37200/IJPR/V24I8/PR280053
Received: 19 Jan 2020 | Revised: 06 Feb 2020 | Accepted: 12 Mar 2020

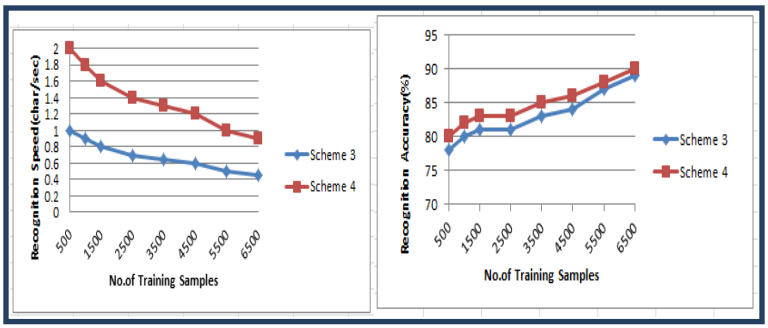

Figure 5: Performance of Recognition speed Figure 6: Performance of Recognition accuracy

V.CONCLUSION AND FUTURE WORK

We have proposed a new feature for online handwritten stroke recognition. Thisfeature is based on direction operator, which represents relative position of each coordinate with respect to other coordinates in the stroke. This feature captures structural, global and local characteristics of a stroke. Hence it gives improved performance over simple (x,y) feature and (x,y) & tangent angle feature. Our proposed Hybrid approach shows that the recognition time and space occupied by proposed feature vector are less as it contains only binary values in addition to improved recognition accuracy. To continue our work, we will check performance of direction operator feature when it is combined with other features by using different classifier combinations in hybrid scheme.

REFERENCES

1. Srilakshmi .I,RajeshwaraRao.R, “Survey on Online Handwritten Indian Character Recognition and its Extension to Telugu Character Recognition”, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-6S5, April 2019
2. Srilakshmi.I,RajeshwaraRao.R,”Preprocessing of online handwritten Telugu character recognition”International Journal of Advanced and Applied Sciences, 4(7) 2017, Pages: 179-189
3. D. Das, R. Devi, S. Prasanna, S. Ghosh and K. Naik "Performance comparison of online handwriting recognition system for assamese language based on hmm and svmmodelling", International Journal of Computer Science & Information Technology, vol. 6, no. 5, 2014 .
4. Mary, a. Anci manon, m. Bhuvaneswari, n. Haritha, v. Krishnaveni, and b. Punithavathisivathanu. "design of automatic number plate recognition system for moving vehicle." international journal of communication and computer technologies 7 (2019), 1-5. Doi:10.31838/ijccts/07.sp01.01
5. Bahlmann, C., 2006. Directional feature in online handwriting recognition. Pattern Recognition 39(1), 115–125.
6. N. Kato, M. Suzuki, S. Omachi, H. Aso, and Y. Nemoto.A handwritten character recognition system using directional element feature and asymmetric Mahalanobis distance. IEEE Trans. on Pattern Analysis and Machine Intelligence, 21(3):258–262, March 1999.
7. Mausam j. Naik (2019) mapk signalling pathway: role in cancer pathogenesis. Journal of Critical Reviews, 6 (3), 1-6. doi:10.22159/jcr.2019v6i3.31778
8. Kherallah, M., Haddad, L., Alimi, A., Mitiche, A., 2008. On-line handwrit ten digit recognition based on trajectory and velocity modeling. Pattern Recognition Letters 29, 580–594.
9. Mori, M., Uchida, S., Sakano, H.: Global feature for online character recognition. Pattern Recognit.Lett. 35, 142–148 (2014)
10. W. Zeng, X. Meng, C. Yang, L. Huang, Feature extraction for online handwritten characters using Delaunay triangulation, Comput. Graph.UK 30 (2006).

DOI: 10.37200/IJPR/V24I8/PR280053
Received: 19 Jan 2020 | Revised: 06 Feb 2020 | Accepted: 12 Mar 2020

11. Mori, M., Wakahara, T., Ogura, K., 1998. Measures for structural and global 334 shape description in handwritten kanji character recognition. In: Document Recognition V. Vol. 3305. pp. 81–89
12. K. H. Aparna, Vidhya Subramanian, M. Kasirajan, G. Vijay Prakash, V. S. Chakravarthy, SriganeshMadhvanath, "Online handwriting recognition for Tamil", Proceedings of the Ninth International Workshop on Frontiers in Handwriting Recognition (IWFHR' 04), Tokyo, Japan, 2004, pp 438-443.
13. SubhasisMandal,S.R,SureshSundaram," An improved discriminative region selection methodology for online handwriting recognition",InInternational Journal of Document Analysis and Recognition ,Volume 22 Issue 1,March 2019,Pages 1-14
14. Pathirage Kamal Perera. "Traditional medicine-based therapies for cancer management." Systematic Reviews in Pharmacy 10.1 (2019), 90-92. Print. doi:10.5530/srp.2019.1.15
15. Singh, S., Sharma, A., Chhabra, I.: A dominant points-based feature extraction approach to recognize online handwritten strokes. Int. J. Doc. Anal. Recognit. (IJDAR),2007, **20**(1), 37–58
16. Mandalapu, D., Krishna, S. M., 2007. A feature based on encoding the relative position of a point in the character for online handwritten character recognition. In: ICDAR'07. Vol. 2. pp. 1014–1017
17. Srilakshmi .I,RajeshwaraRao.R, "Preprocessing of online handwritten Telugu character recognition",International Journal of Control Theory and Applications 8 (5), 1939-45,2015
18. Zhang, W., Liu, H., Al-Shabrawey, M., Caldwell, R., Caldwell, R.Inflammation and diabetic retinal microvascular complications(2011) Journal of Cardiovascular Disease Research, 2 (2), pp. 96-103. DOI: 10.4103/0975-3583.83035

DOI: 10.37200/IJPR/V24I8/PR280053
Received: 19 Jan 2020 | Revised: 06 Feb 2020 | Accepted: 12 Mar 2020