Printed and Handwritten Tamil Characters Recognition Using Fuzzy Technique

R.M. Suresh, *Member IAENG*

Abstract— **In this paper various Handwritten Tamil characters are classified as one among the prototype characters using a feature called distance from the frame and a suitable membership function. The unknown and prototype characters are preprocessed and considered for recognition. The algorithm is tested for about 2500 samples of Tamil characters and the success rate obtained varies from 76% to 94%.**

Index Terms— **Fuzzy Approach, Printed Characters, Handwritten Characters, Recognition**.

I Introduction

An attempt is made to use fuzzy concept as a tool for recognition of handwritten Tamil Characters. The feature vector consists of distances of the pattern from the frame in sixteen different directions. Actually, two sets are considered for recognition, one with segments and the other with membership values. In this paper the first section deals with an introduction to Tamil alphabets and in the second section we present the preprocessing methods and finally the classification and main results.

II Introduction to Tamil Alphabets and other Characters

The alphabets of Tamil are very old and are organized in a systematic way. The alphabet set splits into set of vowels, consonants, composite letters, and special letter. Of course some characters from Sanskrit are also being used in Tamil text. A complete set of the Tamil alphabets is shown in Chinnuswamy & Krishnamoorthy [8]. There are 12 vowels, 18 consonants, 216 composite letters, one special character (AK) and 14 other characters. Composite letters are not basic and they are derived by combining consonants and vowels as described in [9]. We have identified some 67 Tamil characters as the basic characters (Vowels, Consonants, and composite letters) and if one recognizes these 67 characters then all the 247 characters can be recognized. The list of 247 and 67 characters are represented in Table 1 and Table 2.

The author currently works at RMD Engineering College, Kavaraipettai, Tamil Nadu, INDIA. The author thanks the support provided the Management of RMD Engineering College Shri R.S. Munirathinam, Shri R.M. Kishore and Shri. R. Jothi Naidu and the Dr. V.C. Veera Reddy. Principal of the college.

Table 1 Tamil Characters

Vowels அ ஆ இ ஈ உ ஊ எ ஏ ஐ ஒ ஓ ஔ ஃ
Consonants க ங ச ஞ ட ண த ந ப ம ய ர ல வ ழ ள ற ன
I. Compound Characters
க கா கி கீ கு கெ கே கை கொ கோ கௌ க் ங ஙா ஙி ஙீ ங ஙூ ஙெ ஙே ஙை ஙொ ஙோ ஙௌ ங் ச சா சி சீ சு சூ செ சே சை சொ சோ சௌ ச் ஞ ஞா ஞி ஞீ ஞு ஞூ ஞெ ஞே ஞை ஞொ ஞோ ஞௌ ஞ் ட டா டி டீ டு டூ டெ டே டை டொ டோ டௌ ட் ண ணா ணி ணீ ணு ணூ ணெ ணே ணை ணொ ணோ ணௌ ண் த தா தி தீ து தூ தெ தே தை தொ தோ தௌ த் ந நா நி நீ நு நூ நெ நே நை நொ நோ நௌ ந் ப பா பி பீ பு பூ பெ பே பை பொ போ பௌ ப் ம மா மி மீ மு மூ மெ மே மை மொ மோ மௌ ம் ய யா யி யீ யு யூ யெ யே யை யொ யோ யௌ ய் ர ரா ரி ரீ ரு ரூ ரெ ரே ரை ரொ ரோ ரௌ ர் ல லா லி லீ லு லூ லெ லே லை லொ லோ லௌ ல் வ வா வி வீ வு வூ வெ வே வை வொ வோ வௌ வ் ழ ழா ழி ழீ ழு ழூ ழெ ழே ழை ழொ ழோ ழௌ ழ் ள ளா ளி ளீ ளு ளூ ளெ ளே ளை ளொ ளோ ளௌ ள் ற றா றி றீ று றூ றெ றே றை றொ றோ றௌ ற் ன னா னி னீ னு னூ னெ னே னை னொ னோ னௌ ன்
Grantha Characters(Special Letters) ஸ ஷ ஜ ஹ க்ஷ ஶ்ரீ
II. Their Compound Characters
ஸ ஸா ஸி ஸீ ஸு ஸூ ஸெ ஸே ஸை ஸொ ஸோ ஸௌ ஸ் ஷ ஷா ஷி ஷீ ஷு ஷூ ஷெ ஷே ஷை ஷொ ஷோ ஷௌ ஷ் ஜ ஜா ஜி ஜீ ஜு ஜூ ஜெ ஜே ஜை ஜொ ஜோ ஜௌ ஜ் ஹ ஹா ஹி ஹீ ஹு ஹூ ஹெ ஹே ஹை ஹொ ஹோ ஹௌ ஹ் க்ஷ க்ஷா க்ஷி க்ஷீ க்ஷு க்ஷூ க்ஷெ க்ஷே க்ஷை க்ஷொ க்ஷோ க்ஷௌ க்ஷ்

Table 2 Reduced Character Set identified for recognition

Compound Characters கு கூ சு சூ ஞு ஞூ டி டீ டு டூ ணு ணூ து தூ நு நூ மு மூ ரு ரூ ழு ழூ ஞ ஞூ று றூ னு னூ
Glyph Letters ி ீ ெ ே ை ா ் [{] } -

ISBN: 978-988-98671-8-8

III Handwritten and Printed Tamil Characters Recognition

3.1 The Proposed System

The recognition technique reported here consists of two parts. The first part deals with preprocessing of both prototype and unknown characters in a compatible and reliable form. The second part is related to the feature extraction and recognition, where a set of flexible models of prototype character is created. In this part, an unknown character is recognized by finding out the maximum membership value and similarity in number of segments between the unknown input character and prototype class.

3.2 Preprocessing

The preprocessing stage consists the following steps:

1. The handwritten Tamil character is scanned using flat bed scanner. The digitized input is considered to have fixed height (15 pixels) and width (30 pixels).
2. Histogram based threshold approach is used to convert the gray image into two tone image.
3. The edges of the two tone image is obtained
4. Polygonal approximation of a character image is obtained from its skeleton. Sample are shown in Figure 1a, 1b.
5. The pattern primitives are identified (Fuzzy functions), and labeled for the two tone image.
6. Labeling Algorithm is used. [9]

3.3 Feature Extraction and Primitive labeling

The preprocessed input character x is considered and its string is compared with the string of each of seven prototype characters Y. Let Ny be number of segments in x for which match is found in Y. Let

$$\delta y = \frac{Ny}{\text{Total number of segments in input character x}}$$

The two tone converted, edge detected, polygonal handwritten character (TEPHC) is brought into the frame of size 64x64 (refer figure 2). Along each of the 16 directions, the distance from the frame to the point where the direction hits the image is measured. This gives a vector (d1,d2,…,d16). In fact, various frame size was considered to find the optimized frame.
Five such vectors are obtained by placing the TEPHC in five different positions in the frame. The average of these five vectors is called the feature vector and it is denoted by (a1,a2,…,a16). Now let ni = ai/max(ai)i=1,2,...16. The vector (n1,n2,…,n16) is called the normalized feature vector (NFV).

Using the above process NFV is obtained for each of the prototype Tamil character. Let Y denote the set of all NFVs of prototype Tamil characters.

3.4 Classification and Recognition

Let x = (nx1,nx2,…,nx16) be the NFV of an input character.

$$\text{Let } \mu(x,y) = 1 - \delta y \left[\sum_{i=1}^{16} (n_i x - n_i y) \right]^{1/2}$$

where $y \in Y$.

Now x is classified as the prototype character Y for which $\mu(x,y)$ [5] is maximum. In this process, it is assumed that the input character is one of the prototype characters.

IV Results and Conclusion

The algorithm was applied for about 2500 samples for each of the seven chosen Tamil characters given in table 3 and the percentage of successful recognition varies from 76% to 94%. This method is more simple and successful than that of the algorithm presented by Chinnuswamy and Krishnamorthy [8]. The main advantage of this algorithm is that it is simple, and a unique NFV is obtained for each character. In the calculation of NFV, since we have taken the average of five vectors corresponding to the placement of two tone converted Tamil character, the method recognizes character which are even tilted up to an angle of 30 degrees[7,9].

References

[1]. W.Pedrycz, Fuzzy sets in Pattern Recognition: Methodology and Methods, Pattern Recognition vol 23, ½, pp 121-146, 1990.
[2]. I.S.I. Abhuhaiba and P.Ahmed, A Fuzzy Graph theoretic approach to recognize the totally unconstrianed handwritten Numerals, Patt. Reco. Vol 26, No. 9, pp. 1335- 1350, 1993
[3]. S.N. Srihari, Recognition of handwritten and machine printed text for postal address interpretation, Patt. Reco. Letters, Vol 14, pp 291-302, 1993
[4]. T. Wakahara, Toward robust handwritten character recognition, Patt. Reco. Letters, Vol 14, 345-354, 1993.
[5]. R.M.K.Sinha et al., Hybrid Contextual Text Recognition with String Matching, IEEE PAMI, Vol 15, No 9, pp 915-923, 1993.
[6]. Z. Chi et al., Handwritten Numeral Recognition using Self-Organizing maps and Fuzzy Rules, Patt. Reco. Vol 28, No. 1, pp 59-66, 1995.
[7]. J. Cao et al., Recognition of Handwritten Numerals with Multiple Feature and Multistage Classifier, Patt. Reco. Vol 28, No 2, pp 153-160, 1995.
[8]. P.Chinnuswamy and S.G.Krishnamoorthy, "Recognition of Handprinted Tamil Characters", Pattern Recognition Vol. 12, pp141-152, 1980.
[9]. R.M.Suresh, "Fuzzy Technique Based Recognition of Handwritten Characters", LNAI 2955, 297-306, 2006.

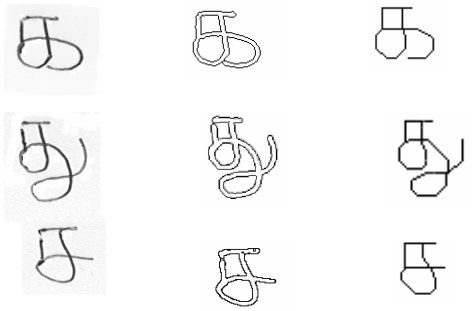

Figure 1 a) Polygonal Approximation for KA, THU & CHA

அகதுசராடமயலஞரமுருநாநிதிதழமீலி

அகதுசராடமயலஞரமுருநாநிதிதழமீலி

அகதுசராடமயலஞரமுருநாநிதிதழமீலி

Figure 1 b) Different presentation of Tamil characters considered for recognition

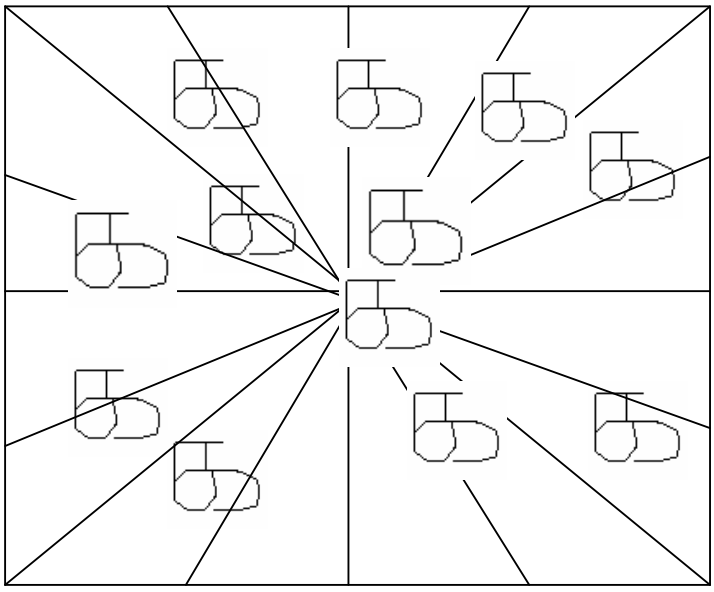

Figure 2. 16 Directions for Feature Extraction.

ISBN: 978-988-98671-8-8

Table 3. Recognition Result of the Handwritten Tamil Characters

Character	String obtained	% of Recognition	% of Error	% of Misc
AA	vvhlhrvlhlvrhlvrhvrhl	88	10	2
KA	hvlhrvhvhvh	88	10	2
THU	hvlhrlvhrvlhrhv	76	20	4
CHA	hvlrvhvh	94	02	4
RAA	vlhrvlhrvlhv	83	12	3
TA	vh	96	04	0
MA	vhvhlvr	94	06	0
YA	vrhvhv	94	06	0
LA	rlhrvhlrlr	83	14	3
GNA	vlrhhvr	86	12	2
RAE	hvv	96	04	0
MU	vhhvhllhrrvlh	88	10	2
RU	hvhvlhrvlr	94	06	0
NAA	hvhrlrlhrlrlhrlrl	78	20	2
NEE	hrvvhlrvrhv	84	14	2
THEE	hvhrvhvhrllvlhrlh	78	20	2
THA	hvhrllvlhrlhhvhrv	85	13	2
ZHA	vlhhvhvrvr	86	12	2
MEE	hvlvhrlhvlvhl	88	12	0
LY	rvhrlrlrrlhrlh	90	09	1

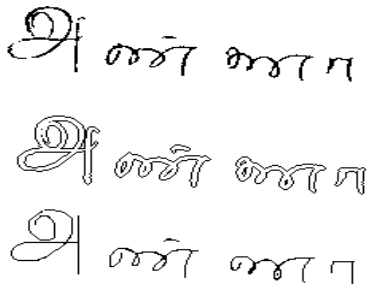

Figure 3. a)

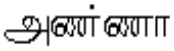

Figure 3. b)

Figure 3 c)

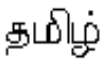

Figure 3 d)

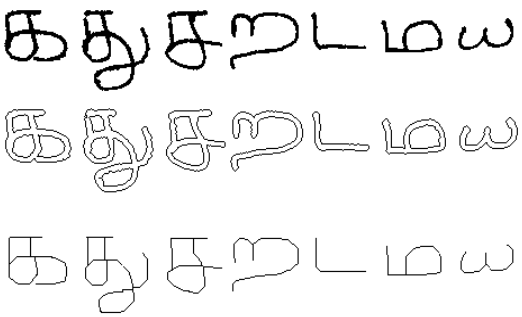

Figure 3. e)

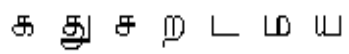

Figure 3. f)

Figure 3 a,c,e) Handwritten characters, Edge, Polygonal image Figure 3 b,d,f) Their Recognized Characters

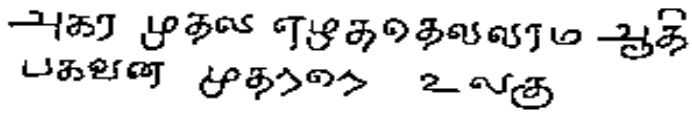

Figure 4. Some of the VATTELUTHU OF 8TH CENTURY considered for Recognition

ISBN: 978-988-98671-8-8