Machine Recognition of Devanagari Hand-printed Script: Character and Word Level Analysis

Satish Kumar
Department of Computer Science and Applications,
Panjab University SSG Regional Centre,
Hoshiarpur, Punjab, India
Email: satishnotra@yahoo.co.in

Abstract—**Various South Asian languages such as Sanskrit, Hindi, Marathi, Nepali etc. are written using Devanagari script and millions of people in world are using this script for written communication. So there is a high demand of good Devanagari ICR but it is still not available. An individual having good knowledge of the script of a language can easily read some words written on a paper pertaining to that script, though these are expressed in very bad manner, on the basis of his/her mental dictionary. Such words may not be easy to read through a machine. A machine-printed word in Devanagari script can be divided into three vertical regions *i.e.* the upper, the lower and the middle. It is difficult to express the size of various regions in hand-printed *i.e.* the size of a region can be bigger or smaller and depends upon a writer's writing style. In case of machine-printed it is easy to recognize a script on the basis of structural features but in case of hand-printed it is difficult due to structural variation in writing. This paper deals with the analysis of Devanagari characters and words according to recognition point of view. The various issues related to the structural features of Devanagari are also discussed. Moreover, the structural features are extracted from skeltonized or contoured image. The various problems likely to be faced while extracting structural features from skeltonized or contoured image are also discussed.**

Keywords- Devanagari; hand-printed; structural, skeleton, contour, character recognition.

1. Introduction

An ICR works in various stages such as scanning, pre-processing, feature extraction, classification and post-processing. Each stage has large number of optional techniques which have been used in different recognition systems. Selecting a best technique for a particular application is a daunting task. One has to exhaustively study the literature, implement them and observe their performance. Obviously this is a big task. The feature extraction method(s) used must be robust for expressing the properties of a character/script under consideration. If there is slight variation in character image either due to printing, writing or due to instrument used, it should be able to absorb the same. There are many feature extraction methods available in literature. The features may be local or global. The features may be extracted from an original character or from its skeleton or contour form. As far as feature extraction stage is concerned, Govindan et al[1] classified the various features in three categories *i.e.* statistical, structural and global transforms and series expansion. Each category has its pros and cons in terms of computational speed, computational complicacy and accuracy. In structural based approach, a character is recognized on the basis of structural primitives from which it is build up and these primitives are also known as character strokes. The number of such stroke components, the kinds of stroke components and the relationship between these components in some order is worked out. Some prominent authors have made prompt efforts to use structural features[6,7,8,13]. The feature extraction process from structural feature is fast [1,2]. Section C covers drawbacks of using structural features.

Statistical Features are based on the statistical distribution of black and white pixels in a character image. These features are invariants to character distortion and writing styles to some extent. In addition to this the feature vector

can be computed using these techniques with high speed. Some most commonly used statistical features for character recognition are zoning, moments, projection histograms, crossings, character loci and *n*-tuple. The feature extraction methods based on profiles, crossings, histograms, zoning have been mostly used as complementary or supporting features to enhance the performance of primary features [14,16-18]. Some authors also have also used a combination of statistical and structural features for handwritten character recognition[3, 5, 9,14]. In addition to above said feature extraction methods, some more features are also available which have very good discrimination ability. Some of these features give very good results and are based on gradient (Sobel or Prewit or Kirsch)[19-25] , chain code histograms of image contour[26-27], distance transform[28-29], pixel distance[30], a fixed size normalized image (binary or gray)[34-35], stroke based[33], feature based on foreground and background information of a character[32], directional distance distribution[29] and directional element feature[31].

Segmentation is one among the various pre-processing steps performed on an image before feature extraction and classification is carried out. The variability caused in script writing is so high that sometimes it becomes difficult even to read a hand-printed material by a well knower of the language and so reading the same script through a machine cannot be expected. In automatic script recognition, a machine is trained to recognize the scripts of a particular language so that it can decide the meaning of a script under study. The process of training a machine is also known as machine learning. In machine recognition process the same strategy is followed. Therefore, there are three ways to automate a system for word recognition: 1) Segmentation Based, 2) Holistic, 3). Hybrid. In segmentation based approach, the given word is segregated into individual components and each component is recognized and assigned a symbol, and the resulting symbols are reassembled to know the identity of a word. Whereas in holistic approach, a word is not broken down into individual components rather it is trained with a complete set of words by extracting their properties. Some segmentation techniques used in Indian languages are given in [36-45].

The various efforts done for the recognition of Devanagari are reported by Ghosh[64] and Pal [65]. The various authors who done work for Devanagari script recognition prior to 1990 are :[47,50-52]. The major studies on machine-printed script of Devanagari after 1990 are:[23-26]. The various studies on Devanagari character recognition are: Pal et al[53], Sharma et al[54], Deshpande et al [55] and Kumar [56-57],. The various studies on hand-printed Devanagari word recognition are: Shaw et al[58], Parui et al[59], Ramteke et al[61], Jayadevan et al[62], Oval[63] and Singh[60].

Hindi is widely used language in south Asian region which is written in Devanagari script. Hindi is also official language of India. This research paper covers the various issues related to the recognition of Devanagari hand-printed characters and words. Section 2 covers Devanagari script analysis. Section 3 covers various structural issues in character recognition. Section 4 covers various structural issues in word recognition. Section 5 covers recognition complexities of Devanagari Words. Section 6 covers issues arising due to skeletonization of word level images. The discussion and conclusion is coved in Section 7.

2. Script Analysis

In case of Devanagari script based languages, the words are written line-by line from top to bottom and left to right like other majority of languages of world. So segmenting a page into lines and then lines into words follow the same strategy as it is used for other languages. But segmentation technique required to segment a word into characters is language-dependent. Among Hindi and Sanskrit which are written using Devanagari script, it is difficult to recognize Sanskrit as its words are longer and composed of a large number of half characters and / or vowels (matra) which are difficult to locate not only in hand-printed but in machine printed too. The words in Devanagari script are not written in cursive manner but there are some alphabets which are written in cursive manner. Further, the style of individual writing decides whether its expression is cursive or not. In machine printed, variation in representation depends upon

the type-set of a font but in case of hand-print there is no limit. The basic Devanagari alphabets generally used in Hindi along their position in a word is given in Table 1.

Table 1: Devanagari Alphabet set along with their position in a word.

Region		**Symbols**
Upper		◌ ◌ ◌ ◌ ◌ ◌ ◌ ◌
Middle	full	अ इ उ ऊ ए ऋ क ख ग ध ड च छ ज झ ण ट ठ ड ढ त्र त थ द ध न प फ ब भ म य र ल व स श ष ह क्ष त्र ज्ञ श्र क ख झ फ ऱ श रु रू द्र ह्म द्य ा :
	Touching with Head-line (Occupy upper part only)	ग ह ण ह ह भ म र
	Non-Touching with Head-line	ट ल झ ज ञ च ढ ट
Lower		ु ृ ८ ं .

Like other languages, in Devanagari script too, a word is vertically divided into three regions: the upper region, the middle region and the lower region. The upper region is occupied by the vowels or the part of vowels; the lower region is also occupied by the vowels while the middle region is occupied by the vowels, consonants, pure consonants, compound characters, and a vowel (matra) or a part of matra. A hand-printed Devanagari word demonstrating all the three vertical regions is given in Fig 1.

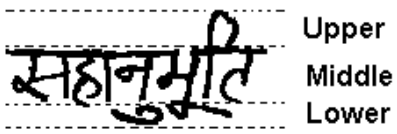

Fig 1: Devanagari word divided into three vertical regions.

Some symbols mentioned as “Non Touching with Head-line” in Table 1 cover whole lower part of middle region where as in some cases lower part of middle region is a little bit empty. The above said fact is true only for the machine–printed. Some full size, head-line touching and non head-line touching characters in some Devanagari words in machine-printed form are depicted in Fig 2.

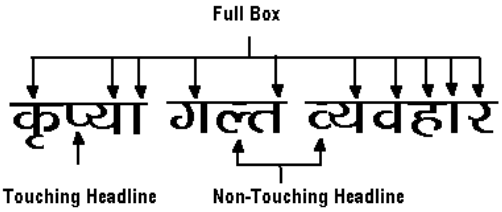

Fig 2: The alphabets of various categories Devanagari words.

3. Structural Issues in Character Recognition

Structural features are based on the geometrical and topological properties of a character and these properties may be local or global [1]. A character is composed of number of components in the form of strokes. These strokes may be lines, arcs, curves, etc. Each character component is called as a character primitive. These are extracted either from skeleton or contour of a character image. The various stroke primitives are extracted and approximated. The relationship between these components is established. The character is recognized on the basis of number of such components, the kinds of components and the relationship between these components in some order. The various structural methods differ in respect of primitive selection and their association for shape depiction. Some commonly used topological and geometrical features in character recognition are: endpoints, T-points, cross points, extrema (top, bottom, left and right), direction of stroke, loops, convex and concave arcs, straight lines, directional points, bend points, etc.

3.1 Drawbacks of Structural Features for Characters Recognition: There are some drawbacks of using structural approaches [5, 16] and these are as:

1). Since the number of primitives present in a character image is not known prior, it is very difficult to detect the primitive and estimate its features. Therefore, the success of applying structural feature depends upon the prior knowledge of shape boundary features stored in a database.

2). The matching schemes used in structural approaches use non-metric similarity measure. These methods do not guarantee best match.

3). The structural features are sensitive to noise and this representation do not preserve the topological structure of an object. With change in boundary of an image the local features change a lot.

4). In structural based approach, though the character primitives provide a stable representation but it does not completely cover the variability in characters. The instability caused in feature space due to incomplete recovery of variable writing styles is not easy to handle in classification stage.

Moreover, the structural features are extracted from a skeleton. Thinning process introduces spurious branches and clusters as given in Fig 3. These defects are common in handwritten character images. The spurious branches and clusters of small size are easy to remove but bigger size poses difficulty.

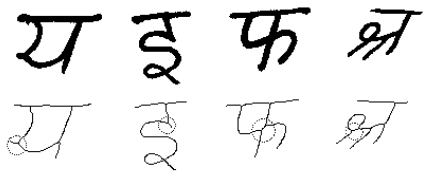

Fig 3: Spurious branches and clusters created due to thinning process.

Even performing primary classification based on topological features in Devanagari handwritten character recognition is difficult. For example, if we want to categorize Devanagari alphabet set into subsets on the basis of number of end points. It is very difficult as the number of end points of a character is different corresponding to different writing. The various intra-class Devanagari handwritten characters with varying number of end points are given in Fig 4. All these characters are taken after head-line removal.

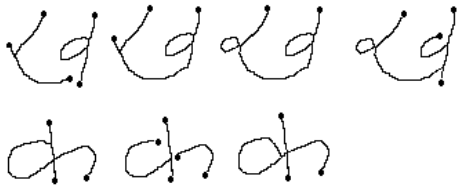

Fig 4: Some Devanagari handwritten characters with varying number of end points.

Some more reasons for not using structural features for Devanagari handwritten character recognition are as follows[46]:

1). Extracting structural features are very difficult

a) Due to complex structure of some of its characters.

b) Intra-touching (mingle) of the various strokes present in a character, Fig 5, due to hasty writing or individual style of writing or ink blot.

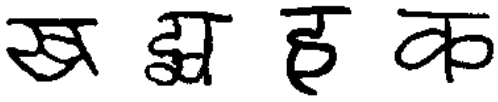

Fig 5: Some Devanagari characters with intra-touching strokes.

c) Some characters are built up from multi-strokes and these primitive may not be touching as it is required in a character, Fig 6, due to hasty writing or individual style of writing.

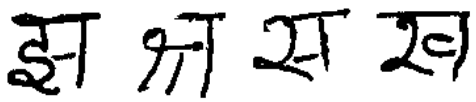

Fig 6: Multi-stroke characters with dissociated primitives.

d) There is blurring present in some characters. Extracting skeleton from blurred image destroy its shape. Fig 7 demonstrates the situation, where encircled parts of the characters are completely vanished.

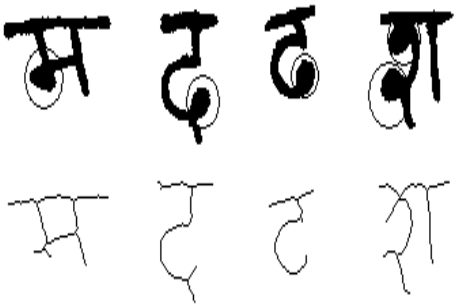

Fig 7: Blurred (encircled) and vanished Devanagari characters

2). It is difficult to extract structural features from noisy images as small loops are created in skeletonization process.

3). Extra branches are created due to fluent writing as shown in Fig 8. Structural features are extracted from the skeletonized image of a character. These extra branches become more apparent after skeltonization of a character image is carried. Such unpredictable behavior is common in skeletonizing algorithms [66].

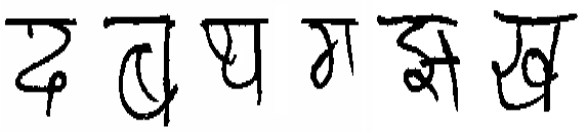

Fig 8: Extra branches produced due to writing.

4. Structural Issues in Word Recognition

In segmentation based approach, a given word is segregated into individual components and each component is recognized and assigned a symbol, and the resulting symbols are reassembled to know the identity of a word. In hand-printed writing the variability caused is so high that it is very difficult to locate boundary between middle and lower regions as the size of different characters varies due to the lack of control while writing. The Fig 9 clearly demonstrates this situation. As a head-line exists in case of words / characters written in Devanagari script, it is easy to locate boundary between upper and middle regions and helpful in segregating / locating the symbols of upper region. It is very difficult to locate the boundary between middle and lower region.

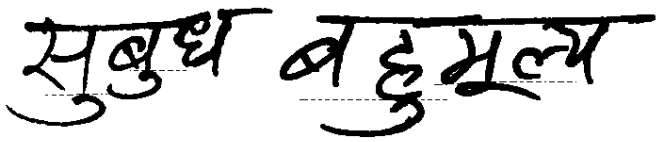

Fig 9: Some full size, head-line touching and non head-line touching characters in some Devanagari words in hand-printed form.

Though some characters of Devanagari script are cursive in writing but while writing a word each character is written individually. A Devanagari word is formed by connecting its letters with each other through head-line. However, a character in a word may touch with adjoining character either due to hasty writing or due to ink stains.

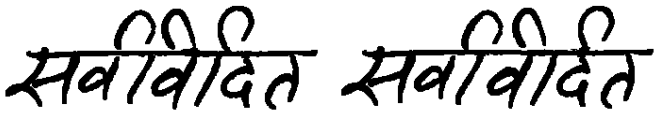

Fig 10: A Hindi word where matra not in its original position(right).

In case of Devanagari, some vowels are formed with the symbols those occupy both upper and middle regions. In recognition process, it is important to decide about a vowel by considering the various parts of a symbol presented both in upper and lower regions which are some what difficult to locate in the sense that they may be written / placed in a word deviating from the rules. The recognition process may produce the output of such image as ambiguous word. The possible word images with exact location of vowels for the word image demonstrated in Fig 10(left). However, with minor change in location of symbols in upper or middle creates a new word as given in Fig 10(right).

In addition to this, some symbols of Devanagari script are built up with multiple strokes which may be dissociated from each other in a given word. In such cases decision about a character must be made keeping in view the various available stroke components. In Fig 11, ख, अ, क, त, ल, न, प, ध are in broken form and there may be difficulty in segmenting and recognizing such characters.

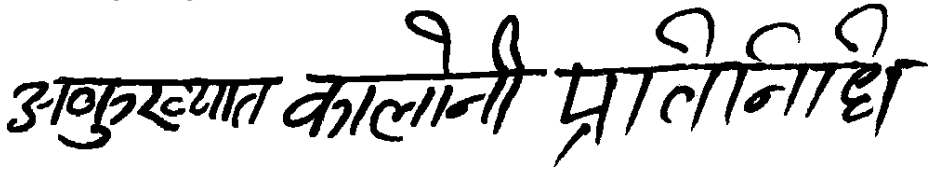

Fig 11: A word written in Devanagari script having broken character primitives

Among the various middle region full size symbols only ‘ ा’ (kanna) and ‘:’ (visarga) belong to vowel or a part of vowel and remaining symbols are either consonants or pure consonants or compound characters. Visarga lies in middle region, Fig 12, may create confusion with noise present in an image. The symbol ‘ ा’ (kanna) is most frequently used vowel in Devanagari where as the use of symbol ‘:’ (visarg) is language dependent.

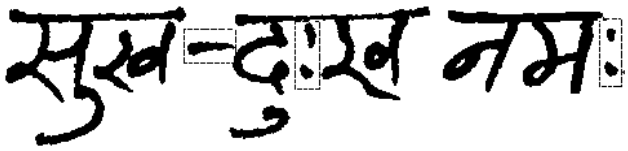

Fig 12: A matra or any other symbol in middle region.

In Hindi it is scarcely used and in Sanskrit it is frequently used. The use of ‘ ा’ and ‘:’ in middle region is demonstrated in Fig 12. The use of visarga is very less in middle of a word where its use at end of word is common in Sanskrit. Like Roman, “-” (hyphen) is also used in Devanagari to join two words. This also lies in middle region but its presence is outside a word. For recognition purpose, it also needs to be treated as a special symbol.

5. Recognition Complexities of Devanagari Words

There are three kinds of approaches to recognize the words of any language *i.e.* segmentation based, segmentation free (holistic) and hybrid. The segmentation free approach, as already mentioned, requires a big lexicon of words which needs a lot of efforts for lexicon creation. The segmentation based approach is easy but it is difficult to carry out segmentation process in it. Some issues required to be dealt with the recognition of words pertaining to Devanagari scripts using this approach are as follows:

1) There are two ways to segregate Devanagari words into characters and then recognize them individually *i.e.* either after removing head-line or without removing head-line. While removing head-line some parts of middle box characters\matra chopped off due to which it poses a problem to recognize various individual characters present in a word. How the head-line from a word may be removed without chopping off important parts of various characters in a given word image?

2) If a word is segmented without removing head-line, it is difficult to estimate the left and right boundary of a character in a word which is non-linear. Also it becomes difficult to decide the dissociation point between two characters even if two adjoining characters are not fused.

3) Due to hasty writing, the two adjoining characters may touch each other. How to locate the boundary between touching characters?

4) Some multi-stroke characters, Fig 11, in a word may not be connected to each other as these should be. How those components are recognized to take decision about the exact/ tentative character?

5) There may be a vowel attached to a character on its lower part. Fig 9 demonstrates the situation. How to decide the presence/ absence of a vowel on the lower part of a character?

6) There are some vowels in Devanagari, Fig 10(right), which may be mis-positioned both in upper or middle regions. How to locate and recognize such vowels?

7) Some pure consonants (half characters) presented inside a word may form a complex structure by combining with other characters. Some such complex structures formed due to the merging of a consonant and a pure consonant are demonstrated inside dotted rectangle areas in Fig 13. How to locate and recognize such components?

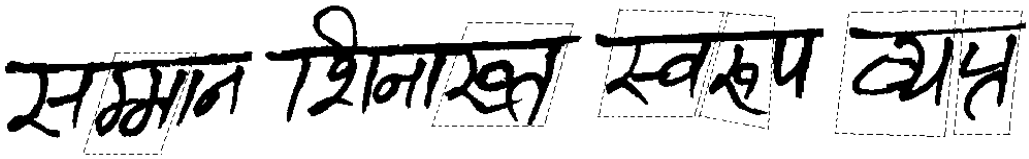

Fig 13: Some complex structures formed due to the merging of consonants and pure consonants.

8) Various matra attached on lower or upper part of a character may be bigger in size due to which it encroach the areas of the adjoining characters. In machine printed the size of such matra or a part of matra remains equivalent to characters size to which it is attached. Such matra may create confusion about its location. This aggregates the ambiguity about final word recognition. Fig 14 depicts the situation.

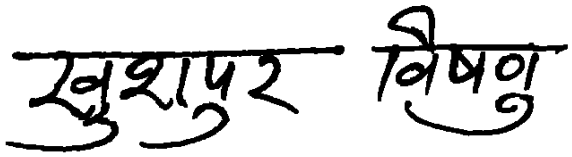

Fig 14: Confused location of lower Matra.

9) Various complex compound characters are formed due to merge of two or more characters. Since, the characters are written left to right, the fusion is expected horizontal as given in Fig 14(a), inside dotted circle, but this is not only the case in Devanagari where the fusion also occurs vertically as given in Fig 14(a) , inside dotted rectangle, making it difficult to know whether to dissociate such fused compound characters horizontally or vertically.

Fig 14 (*a*): Horizontal and vertically fused characters.

10) Some compound characters, built up due to combine of two characters, are also represented using special symbol in Devanagari as given in Fig 14(b). These symbols form a complex structure and are a source of complexity for recognition. The numbers of such characters are more in Sanskrit than Hindi.

Fig 14 (*b*): Special symbol for fused characters.

11) Some awkward styles of writing such as undeveloped, unusual and extra stroke writing also pose problems in Devanagari word recognition. In addition some irregularities committed by the writers while writing further adds complicacies in Devanagari word recognition problem[67].

6. Word Level Skeletonization or Contouring Issues

A word is composed of number of characters each of which is formed from a number of stroke primitives. These strokes may be lines, arcs, curves, etc and may or may not be connected to each other depending upon the structure of a character. The stroke primitives can be extracted from either skeleton or contour of a character image. In structural based recognition process, the various stroke primitives of a character are extracted and approximated. The relationships between various stroke components are established. Some difficulties in recognizing words pertaining to Devanagari script using structural approach which is based on skeletonized or contoured image of a word are as follows:

1). A Devanagari word consists of a head-line. When the given word image is skeletonized, the resultant head-line produced is not smooth which is difficult to remove. One such irregular structure produced in head-line is given in Fig 15 mentioned using dotted rectangle area.

2). When a word is segmented after skeletonization, the two characters which are fused may form a cluster or multiple joint which is difficult to segment. The situation is given in Fig 15 mentioned using dotted circles.

3). In segmentation based approach, each segmented character/ part of character is matched against some already known characters. If a character is in skeleton form, then the skeletonization process may have lost important part of a character image or may have created extra parts in a character image or may have created clusters in a character image etc., which complicates the recognition process a lot.

Fig 15: Skeletonized image of hand-printed Devanagari word image.

4). The various character level recognition problems based on skeletonized character images mentioned in Section 3.1 further adds to word level recognition.

5). During recognition process, the character images are normalized so that proper mapping between stored image and tested images is done. But it is difficult to decide that whether first skeletonization is performed or normalization.

6). On the other hand, if a character\word structure is compared from the contoured image then it has its own problems. The contoured image of a word, Fig 9, is given in Fig 16. An unwanted structure (lump) is create on boundary where extra stokes of a character are generated or ink blot occurs on boundary during writing process.

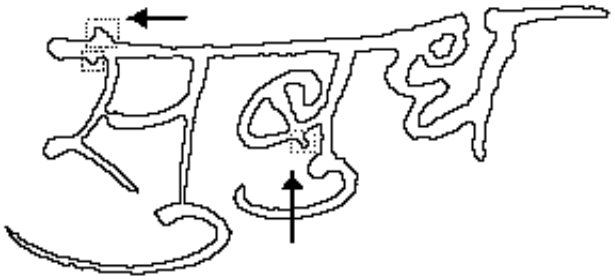

Fig 16: The contoured image of a word given in Fig 9.

7). Some unwanted complex structures are formed on fusion point of two or more characters, Fig 17 doted circular area, due to which segmentation or comparison becomes difficult.

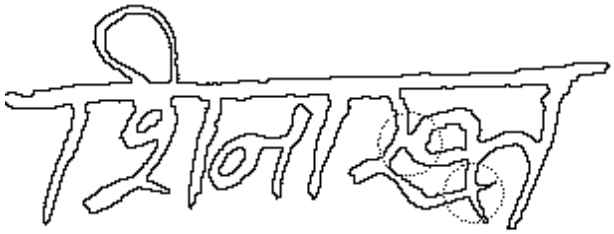

Fig 17: The contoured image of a word given in Fig 13.

8). Moreover, if character\word strokes do not contain adequate size due to poor scanning, thin writing instrument tip, resizing or other reason, the contoured image of such character\word produce a skeleton as given in Fig 18. The resultant image becomes the mixer of contour and skeleton. This also poses difficulty in comparison.

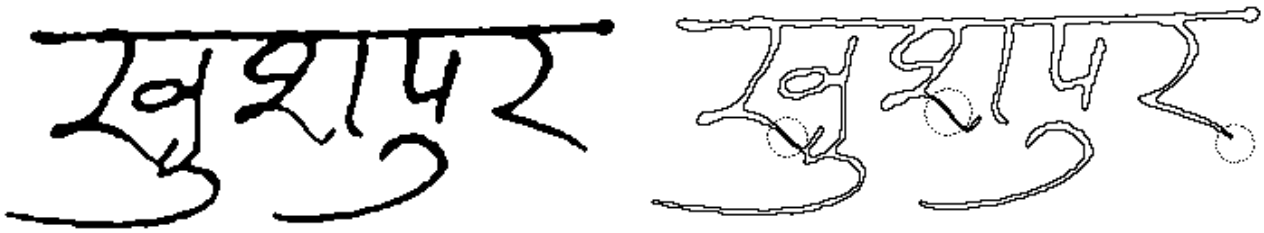

Fig 18: The skeleton strokes produced, dotted circles, during contouring.

9). The images resized with normalization process also, sometimes, destroys strokes present in a character/word image due to which it becomes difficult to extract contour.

Some above mentioned issues are not only hindrance in Devanagari based script recognition but are ubiquitous to all scripts.

7. Discussion and Conclusion

Though a lot of work has been done for the development of recognition system for the various languages of the world but good ICR of majority of languages are still not available. Among the various Indian scripts, Devanagari is such script. The words of Devanagari script are formed by connecting various characters using head-line. Due to the presence of head-line it is easy to separate the symbols lying between upper region and middle region. However, it becomes difficult to separate the symbols between middle and lower region due to absence of any such line/ boundary. Though its words are not cursively written but some its characters are cursive in writing. Devanagari script has its own kinds of problems in respect of recognition. We have studied the problems associated with the recognition of Devanagari hand-printed based on structural features. Also, it is difficult to recognize the hand-printed script using structural features which are extracted from skeltonized or contoured character/word images. The various issues related to structural features also discussed. Some issues related to segmentation of hand-printed Devanagari are also discussed.

References

[1] V. K. Govindan and A. P. Shivaprasad, "Character Recognition – a Review", Pattern Recognition, Vol. 23, No. 7 (1990).

[2] M. S. Khorsheed, "Off-line Arabic Character Recognition - a Review", Pattern Analysis and Applications, Vol. 5, pp. 31-45 (2002).

[3] H. S. Baird, "Feature Identification for Hybrid Structural/Statistical Pattern Classification", Computer Vision, Graphics and Image Processing, Vol. 42, pp. 318-333(1988).

[4] K. Anisimovich, V. Rybkin, A. Shamis and V. Tereshchenko , "Using Combination of Structural, Feature and Raster Classifiers for Recognition of Hand-printed Characters", Proceedings of Fourth International Conference of Document Analysis and Recognition, Ulm, Germany, pp. 881-885 (1997).

[5] P. Foggia, C. Sansone, F. Tortorella and M. Vento, "Combining Statistical and Structural Approaches for Handwritten Character Description", Image and Vision Computing, Vol. 17, pp. 701-711(1999).

[6] X. Li, W. Oh, J. Hong and W. Gao, **"**Recognizing Components of Handwritten Characters by Attributed Relational Graphs with Stable Features", Proceedings of the International Conference on Document Analysis and Recognition, Ulm, Germany, pp. 616-620(1997).

[7] J. Rocha and T. Puvlidis, "A Shape Analysis Model with Applications to a Character Recognition System", IEEE Transactions on Pattern Analysis and Machine Intelligence**,** Vol. 16, No. 4, pp. 393-405 (1994).

[8] K. T. Miura, R. Sato and S. Mori, "A Method of Extracting Curvature Features and Its Applications to Handwritten Character Recognition", Proceedings of the International Conference on Document Analysis and Recognition, Ulm, Germany, pp. 450-454(1997).

[9] J. Cai and Z-Q Liu, "Integration of Structural and Statistical Information For Unconstrained Handwritten Numeral Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 21, pp. 263-270(1999).

[10] G. S. Lehal and Chandan Singh**, "**A Gurmukhi Script Recognition System", International Conference on Pattern Recognition, Barcelona, Spain, Vol. 2, pp. 557-560(2000).

[11] B. B. Chaudhuri and U. Pal, "A Complete Printed Bangla OCR System", Pattern Recognition, Vol. 31, No. 5, pp. 531-549 (1998).

[12] R. C. Gonzalez and R. E. Woods, "Digital Image Processing", 2nd Ed. 2002, Pearson Education.

[13] H. Xue and V.Govindaraju, "Building Skeletal Graphs for Structural Feature Extraction on Handwriting Images", Proceedings of the Sixth International Conference on Document Analysis and Recognition, Seattle, WA, USA, pp. 96-100(2001).

[14] L. Heutte, T. Paquet, J. Moreau, Y. Lecourtier and C. Olivier, "A Structural / Statistical Feature Based Vector for Handwritten Character Recognition", Pattern Recognition Letter, Vol. 19, pp. 629-641(1998).

[15] D. Zhang and G. Lu, "Review of Shape Representation and Description Techniques", Pattern Recognition, Vol. 37, pp. 1-19 (2004).

[16] N. Arica and F. T. Yarman-Vural, "Optical Character Recognition for Cursive Handwriting", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 6(2002).

[17] K. M. Kim, J.J. Park, Y.G. Song, I. C. Kim and C. Y. Suen, "Recognition of Handwritten Numerals Using a Combined Classifier with Hybrid Features", SSPR & SPR, LNCS 3138, pp. 992-1000(2004).

[18] J. T. Favata, G. Srikantan and S. N. Srihari, "Handprinted Character / Digit Recognition using a Multiple Feature/Resolution Philosophy", Proceedings of the International Workshop on Frontiers of Handwriting Recognition", pp. 57-66(1994).

[19] K. Roy, T. Pal, U. Pal and F. Kimura, "Oriya Handwritten Numeral Recognition System", Proceedings of the Eight International Conference on Document Analysis and Recognition, Vol. 2, pp. 770- 774 (2005).

[20] G. Srikantan, S. W. Lam and S.N. Srihari, "Gradient-Based Contour Encoding for Character Recognition", Pattern Recognition, Vol. 29, No. 7, pp. 1147-1160 (1996).

[21] H. Liu and X. Ding, "Handwritten Character Recognition using Gradient Feature and Quadratic Classifier with Multiple Discrimination Schemes", Proceedings of the Eighth International Conference on Document Analysis and Recognition, pp. 19-25(2005).

[22] H. Fujisawa and C.–L. Liu, "Directional Pattern Matching for Character Recognition Revisited", Proceedings of the Seventh International Conference on Document Analysis and Recognition, pp. 794-798(2003).

[23] S.-B. Cho, "Neural-Network Classifiers for Recognizing Totally Unconstrained Handwritten Numerals", IEEE Transactions on Neural Networks, Vol. 4, No. 1, pp. 43-53(1997).

[24] K. M. Kim, J.J. Park, Y.G. Song, I. C. Kim and C. Y. Suen, "Recognition of Handwritten Numerals Using a Combined Classifier with Hybrid Features", SSPR & SPR, LNCS 3138, pp. 992-1000(2004).

[25] S. Knerr, L. Personnaz and G. Dreyfus, "Handwritten Digit Recognition by Neural Networks with Single-Layer Training", IEEE Transactions on Neural Networks, Vol. 3, No. 6, pp. 962-968(1992).

[26] F. Kimura and M. Shridhar, Handwritten Numeral Recognition Based on Multiple Algorithms, Pattern Recognition, Vol. 24, No. 10, pp. 969-983(1991).

[27] J. Cao, M. Ahmadi and M. Shridhar, Recognition of Handwritten Numerals with Multiple Feature and Multistage Classifier, Pattern Recognition, Vol. 28, No. 2, pp. 153-160(1995).

[28] Zs. M. Kovics and R. Guerrieri, "Massively-Parallel Handwritten Character Recognition Based on the Distance Transform", Pattern Recognition, Vol. 28, No. 3, pp. 293-301(1995).

[29] II-S. Oh, C.Y. Suen, "Distance Features for Neural Network-based Recognition of Handwritten Characters", International Journal on Document Analysis and Recognition, Vol.1, pp. 73-88(1998).

[30] N. W. Strathy and C. Y. Suen, "A New System for Reading Handwritten ZIP Codes", Proceedings of the International Conference on Document Analysis and Recognition, Montreal, Canada, pp. 74-77(1995).

[31] 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 Transactions on Pattern Recognition Analysis and machine Intelligence, Vol. 21, No. 3(1999).

[32] S. Britto Jr., R. Sabourin, F. Bortolozzi and C. Y. Suen, "Foreground and Background Information in an HMM-based Method for Recognition of Isolated Characters and Numeral Strings", Ninth International Workshop on Frontiers in Handwriting Recognition, Kokubunji, Tokyo, Japan, pp. 371-376(2004).

[33] T. W. Bhowmik, U. Bhattacharya and S.K. Parui, "Recognition of Bangla Handwritten Characters using an MLP Classifier Based on Stroke Features", Proceedings of ICONIP, Kolkata, India, pp. 814-819(2004).

[34] K. Roy, S. Vajda, U. Pal and B.B. Chaudhuri, " A System Towards Indian Postal Automation", Proceedings of the Ninth International Workshop on Frontiers in Handwriting Recognition, PP. 580 - 585 , 2004.

[35] Y. Le Cun, O. Mattan, B. Boser, J.S. Denker et al, "Handwritten Zip Code Recognition with Multilayer Networks", Proceedings of International Conference on Pattern Recognition, Atlantic City, NJ, USA, Vol. 2, pp. 35-40(1990).

[36] V. Bansal, "Integrating Knowledge Sources in Devanagari Text Recognition System", Ph. D Thesis, 1996.

[37] B. B. Chaudhuri and U. Pal, "A Complete Printed Bangla OCR System", Pattern Recognition, Vol. 31, No. 5, pp. 531-549,1998.

[38] B. B. Chaudhuri and U. Pal, "An OCR System to Read Two Indian Language Scripts: Bangla and Devanagari (Hindi)", Proceedings of the Fourth International Conference on Document Analysis and Recognition, pp. 1011–1016,1997.

[39] G. S. Lehal and C. Singh, "Feature Extraction and Classification for OCR of Gurmukhi Script", Vivek, Vol. 12, pp. 2–12,1999.

[40] B. B. Chaudhuri, U. Pal and M. Mitra, "Automatic Recognition of Printed Oriya Script", Sadhana, Vol. 27, pp. 23-34,2002.

[41] T. V. Ashwin and P S Sastry, "A Font and Size-independent OCR System for Printed Kannada Documents using Support Vector Machines", Sadhana, Vol. 27, Part 1, pp. 35–58,2002.

[42] S. Sural and P.K. Das, "An MLP using Hough Transform Based Fuzzy Feature Extraction for Bengali Script Recognition", Pattern Recognition Letter, Vol. 20, pp. 771–782,1999.

[43] U. Garain and B. B. Chaudhuri, "Segmentation of Touching Characters in Printed Devanagari and Bangla Scripts using Fuzzy Multifactorial Analysis", IEEE Transactions on Systems, Man, Cybernatics- Part C: Applications and Reviews, Vol. 32, No. 4,2002.

[44] V. Bansal and R. M. K. Sinha, "Segmentation of Touching and Fused Devanagari Characters", Pattern Recognition, Vol. 32, pp. 875-893 (2002).

[45] A. Bishnu and B. B. Chaudhuri, "Segmentation of Bangla Handwritten Text into Characters by Recursive Contour Following", Proceedings of Fifth International Conference on Document Analysis and Recognition, Bangalore, India, pp. 402-405,1999.

[46] S. Kumar, "Theoretical Analysis of Devanagari Handprinted Characters—A Recognition Perspective", International Journal of Computational Intelligence Research, Volume 6, Number 2, pp. 185–193, 2010.

[47] R. Bajaj , L. Dey and S. Chaudhury, "Devanagari numeral recognition by combining decision of multiple connectionist classifiers", Sadhna, vol. 27, Part 1, pp. 59-72, 2002.

[48] U. Pal and B. B. Chaudhuri, "Printed Devanagari script OCR system", Vivek, vol. 10, pp. 12-24, 1997.

[49] R. M. K. Sinha and H.N. Mahabala "Machine recognition of Devanagari script", IEEE transactions on systems, man, and cybernetics, vol. SHC-9, no.8, pp. 435- 441, 1979.

[50] K. Jayanthit, A. Suzukit , H. Kanait , Y. Kawasoej, M. Kimurat and K. Kido, "Devanagari character recognition using structure analysis", Fourth IEEE region 10th international conference, Bombay, India, pp. 363-366, 1989.

[51] I. K. Sethi and B. Chatterjee, Machine recognition of constrained hand-printed Devanagari, Pattern Recognition, vol. 9, no. 2, pp. 69-75, 1977.

[52] S. D. Connel, R.M.K. Sinha and A. K. Jain, "Recognition of unconstrained on-Line Devanagari characters", Proceedings of the international conference on pattern recognition, Barcelona, Spain, vol. 2, pp. 368-371, 2000.

[53] U. Pal, N. Sharma, T. Wakabayashi, F. Kimura, "Off-line handwritten character recognition of Devanagari script", Proceedings of 9th international conference on document analysis and recognition , vol. 1, pp. 496-500, 2007.

[54] N. Sharma, U. Pal et al ,"Recognition of off-line handwritten Devanagari characters using quadratic classifier", *ICVGIP 2006*, LNCS 4338, pp. 805-816, 2006.

[55] P. S. Deshpande, L. Malik and S. Arora, "Fine classification & recogntiton of handwritten Devanagari characters with regular expression & minimum edit distance method", Journal of computers, vol. 3, no. 5, pp. 11-17, May 2008.

[56] S. Kumar, "Devanagari Hand-printed Character Recognition using Multiple Features and Multi-stage Classifier", International Journal of Computer Information Systems and Industrial Management Applications , Vol.2, pp.039-055, 2010.

[57] S. Kumar, "Performance comparison of features on Devanagari hand-printed dataset", International journal of recent trends in engineering, vol. 1, no. 2, pp. 33-37, 2009.

[58] B. Shaw, S. K. Parui and M. Shridhar, "Off-line handwritten Devanagari word recognition: A segmentation based approach", IEEE , 2008.

[59] S. K. Parui and B. Shaw, "Off-line Devanagari handwritten word recognition: An HMM based approach", Proceedings of PReMI-2007(Springer), LNCS-4815, pp. 528-535, 2007.

[60] B. Singh , A. Mittal, M.A. Ansari and D. Ghosh, Handwritten Devanagari Word Recognition: A, Handwritten Devanagari Word Recognition: A Curvelet Transform Based Approach, International Journal of Computer Science and Engineering, Vol. 3, No. 4,pp. 1658-1665, 2011.

[61] A. S R amteke, andM. E Rane, Offline Handwritten Devanagari Script Segmentation, International Journal Of Scientific & Technology Research Volume 1, Issue 4, MAY 2012

[62] R. Jayadevan , S. R. Kolhe , P. M. Patil and Umapada Pal, Database Development and Recognition of Handwritten Devanagari Legal Amount Words, International Conference on Document Analysis and Recognition, pp. 304-308, 2011.

[63] S.G. Oval and S. Shirawale, Recognizing Devanagri Words using Recurrent Neural Network, Proceedings of the 3rd International Conference on Frontiers of Intelligence Computing: Theory and Applications, pp. 413-421, 2014.

[64] D. Ghosh, T. Dube and A. P. Shivaprasad, Script Recognition – A Review", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol 32, No. 12, pp 2142-2161, 2010.

[65] U. Pal, B.B. Chaudhuri, "Indian Script Character Recognition: A Survey", Pattern Recognition, Vol. 37, pp. 1887-1899, 2004.

[66] R. C. Gonzalez and R. E. Woods, "Digital Image Processing", 2nd Ed., Pearson Education.

[67] S. Kumar, "An Analysis of Irregularities in Devanagari Script Writing – A Machine Recognition Perspective", International Journal on Computer Science and Engineering Vol. 2, No. 2, pp. 274-279, 2010.