# Online Handwriting Recognition of Gurumukhi and Devanagiri Characters in Mobile Phone Devices

Anuj Sharma
Department of Mathematics
Panjab University, Chandigarh
India

Kalpana Dahiya
UIET
Panjab University, Chandigarh
India

## ABSTRACT

This paper presents a system to recognize online handwritten Gurmukhi and Devanagiri characters in touch screen based mobile phones. We have used small line segments (derived from elastic matching and chain code techniques) to recognize Gurmukhi and Devanagiri characters. Mobile phones offer main challenges as: less memory and slow processer speed in comparison to Desktop or notebooks or Tablet PCs. We have proposed a system to work effectively in mobile phones in view to see memory and processor limitations. In Gurmukhi, we have achieved an overall recognition rate as 94.69% for a training set of 5330 characters and test set include 1640 characters; the recognition rate for Devanagiri characters for training set of 1050 characters and test set of 504 characters is 86.90%.

## General Terms

Pattern Recognition, Online Handwriting Recognition,

## Keywords

Online handwriting recognition, Preprocessing, elastic matching, chain code technique, post-processing.

## 1. INTRODUCTION

The portability of mobile phones as compared to computers has proved its one of the useful feature in present digital world. The ease with which we can exchange information between user and mobile phones is of great importance today because input devices such as keypad have limitations vis-à-vis input through natural handwriting. The mobile phones offer some major challenges where applications developed for Tablet PCs or notebooks or desktop based computers cannot be implemented. Some of the common challenges are listed in next paragraph.

The speed of processor in mobile phones is slow in comparison to desktop computers or similar devices and resists to run application with large computations in mobile phones; Memory in mobile phones is small, therefore, large datasets cannot be used; Although mobile phones are very common but most of the mobile phones do not include regional language fonts and also some common mobile phones do not accept fonts other than available in its firmware, this may be improved with more development of mobile phone technologies in near future but till then, handwriting recognition in regional languages is a challenge; Most of the programming tools for mobile phones are different from computers and such tools offer challenges during smooth implementation of algorithms; Performance of emulators of mobile phone programming tools cannot be accepted unless it is implemented in real device as mobile phone. It has been noticed that application execution is slow in mobile phones as compared to its emulator's environment. Also, the surface area to draw handwritten text is comparatively small as compare to cross pad or touch screen based computers.

Despite the shortcomings or challenges in mobiles phones, common features of mobile phones attract large population to use it as minicomputer which is available all the time. Also, mobile phone offers an ideal combination of speech and handwriting. Nowadays, most of the mobile phones are touch screen based and able to run online handwriting recognition applications mainly for English language and use of Asian scripts in mobile phones is not common yet. The present study has been done using Gurmukhi and Devanagiri scripts which are scripts of Punjabi and Hindi languages. These languages are widely spoken across the globe. In this paper, author has explained own developed recognition system for online isolated handwritten Gurmukhi and Devanagiri characters in mobile phone devices. The common Gurmukhi and Devanagiri characters are shown in Figures 1 and 2.

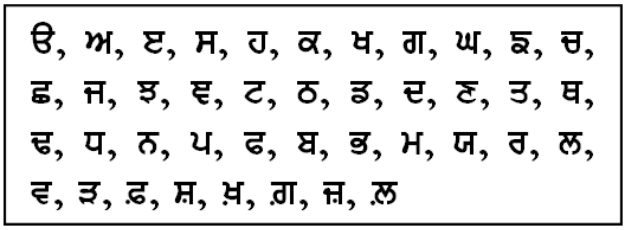


**Fig 1: Common Gurmukhi characters**

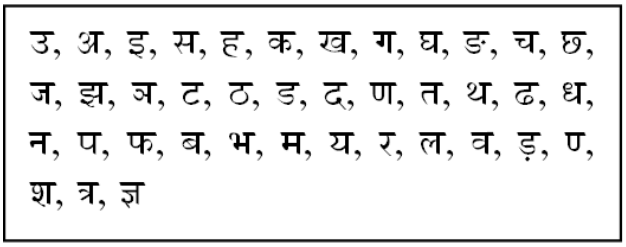


**Fig 2: Common Devanagiri characters**

## 2. SELECTED LITERATURE

The Online handwriting recognition in context of mobile phones is new and literature is available for desktop or tablet pc or notebooks based systems. The selected literature for Gurmukhi, Devangiri and other similar Asian scripts is worth to discuss here in context of similar nature of such scripts. In online handwritten Gurmukhi character recognition, Sharma *et al.* have discussed preprocessing and feature extraction of online handwritten Gurmukhi characters [1]. Sharma *et al.* have recognized online handwritten Gurmukhi characters

using elastic matching and achieved 90.08% accuracy [2]. This study presents grouping of strokes and further recognition of characters in writer independent environment. In recognition of Gurmukhi words, a new stage as rearrangement of strokes after post-processing have been presented by Sharma *et al.* and achieved 81.02% recognition accuracy [3]. In this paper, they have explained nature of dependent strokes and discussed various set of rules associated with these dependent strokes. Sharma and Lehal worked on Gurmukhi script and developed an approach to segment Gurmukhi words using Structural features technique. They achieved overall segmentation as 96.22% for 1907 words [4]. Sachen *et al.* worked on online handwritten Gurmukhi script and presented a method to segment based on pressure, pen down and time parameters [5]. Joshi et al. achieved 94.49% recognition rate using structural recognition and feature based matching for Devanagiri in writer dependent systems[6]. Bhattacharya *et al.* achieved 92.83% recognition rate using HMMs and neural networks for Devanagiri in writer independent system [7]. Swethalakshmi *et al.* achieved 96.69% (42 classes) and 97.27% (82 classes) recognition rate using support vector machine for Devanagiri [8]. Bhattacharya *et al.* used direction code based feature extraction to recognize Bengala characters and achieved recognition rate as 93.90% [9]. Babu *et al.* used HMMs to recognize Telugu symbols and achieved recognition rate as 91.6% [10]. An accuracy of recognition rate with 90.6% has been achieved by Prasanth *et al.* for Telugu symbols using dynamic time warping [11].

## 3. RECOGNITION OF ONLINE HANDWRITTEN GURMUKHI AND DEVANAGIRI CHARACTERS

The established procedure to recognize online handwriting includes: collection of input handwritten stroke; segmentation; preprocessing; computation of features; recognition; post-processing [12]. The present study includes small line segments as recognition method [13-14]. The small line segment method has been derived from elastic matching and chain code techniques.

In present study, the necessary steps to recognize characters have been given below and following paragraphs explain their working:

(i) Input handwritten stroke collection
(ii) Preprocessing of input handwritten characters
(iii) Feature extraction
(iv) recognition
(v) Post-processing to refine results

A character is group of one or many strokes. An input handwritten stroke with pen tablet device captures number of points between single pen-down and pen-up actions. These points are group of co-ordinates and stored sequentially. After pen-up action, collected handwritten stroke is sent for preprocessing. Online handwriting recognition works with handwritten data that is inputted with a digitizing tablet and special stylus. This data is a sequence of (x, y) coordinates. An (x, y) coordinate is called a point and a point together with its stroke number is defined as a node. Initially, we store the handwritten data in a dynamic list that consists of nodes. Structure of a node is given as:

```
struct node
{
        Point P ;
        int StrokeNumber;
};
```

Thus, a stroke is the collection of points in sequential order that are captured between pen down and pen up actions. A character is the collection of one or more strokes. All captured points are stored in sequential order so that directions of pen movement can also be realized. Handwritten characters are stored in a .txt or .xml format so that original points could be used in post processing, comparison of results and further recognition process.

The preprocessing stages used in present study include: (a) size normalization and centering, (b) interpolating missing points, (c) smoothing and slant correction, (d) resampling of points. The details of these preprocessing steps are available in literature [15-20]. Figure 3 presents an input handwritten stroke during preprocessing stages. The features are computed as low-level and high-level features. The low-level features are derived in a stroke whereas high-level features are derived with neighboring strokes. In present study, low-level features are direction, slant, position, aspect ratio, size, area, curliness and linearity, whereas high-level features are loops, headline, straight line, crossings and dots. The high-level features as loops, headline, straight line and crossings have been presented for Devanagiri character 'ख' in figure 4. The details of preprocessing and feature extraction stages in context of Gurmukhi characters have been discussed in previous study [1,14].

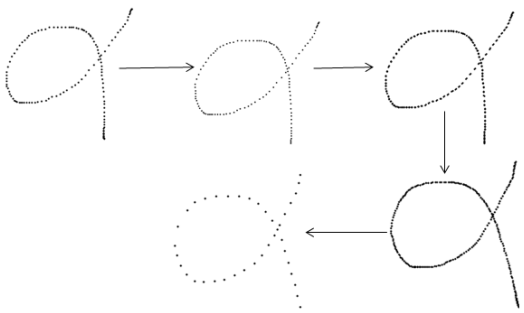

**Fig 3: Preprocessing stages of an input handwritten stroke**

The small line segment as recognition method has been used as it uses less memory and computations [14]. This is possible due to data format and calculations formulas in small line segment. The mobile phone devices need compact data format and also, processor performance does not allow large calculations with large data sets. Therefore, small line segments selection at recognition stage enhanced the performance of developed recognizer. To recognize a character, the recognizer recognizes strokes in sequential order and find their suitable combination to reveal name of character. The strokes prototypes and their procedure of combination of strokes have been discussed in literature using Gurmukhi characters [2-3]. The identification of unique strokes has been done using K-means clustering technique [21]. The K-means clustering common steps are: (1) Place K points into the space represented by the strokes that are being clustered. These points represent initial group centroids; (2) Assign each stroke to the group that has the closest centroid; (3) When all strokes have been assigned, recalculate the positions of the K centroids; (4) Repeat (2) and (3) steps until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

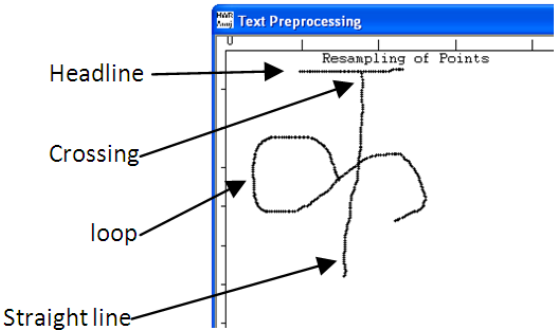


**Fig 4: Common high-level features in Devanagiri character 'k'**

The Figure 5 explains formation of small line segments for a handwritten stroke and these small line segments are given names subject to their angular values as presented in Table 1. We have used fixed number of small line segments as we get fixed number of points from resampling of points step in preprocessing stage in this study. It was done in view to see the mobile phone devices capacity to handle limited number of points in a stroke. The large numbers of points demand heavy calculations and memory storage.

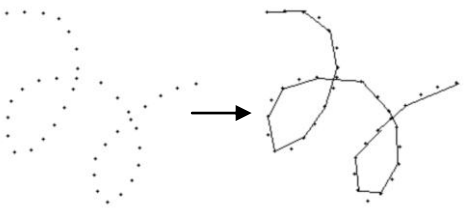

**Fig 5: Small line segments for points of a stroke**

The small line segments vector formed for each stroke is compared against all training data strokes in same format as test data and the stroke with minimum value is selected as recognized stroke class. The detailed explanations of these computations are available in literature [13-14]. In brief, this computation of minimum value is computed using Algorithm 1. As suggested in literature, similar shape characters need special attention. To overcome such issues, post-processing step has been applied after recognition stage in order to refine the recognition results. Common high level features as loop, crossing, presence of headline or vertical line strokes and dots have been used to distinguish such characters in post-processing step. The details of post-processing stage have been discussed in literature [14].

**Table 1. Direction names and their ranges**

| Direction Name | Direction Range |
|---|---|
| A | angle>$345^0$ OR angle≤$15^0$ |
| B | $15^0$ < angle ≤ $45^0$ |
| C | $45^0$ < angle ≤ $75^0$ |
| D | $75^0$ < angle ≤ $105^0$ |
| E | $105^0$ < angle ≤ $135^0$ |
| F | $135^0$ < angle ≤ $165^0$ |
| G | $165^0$ < angle ≤ $195^0$ |
| H | $195^0$ < angle ≤ $225^0$ |
| I | $225^0$ < angle ≤ $255^0$ |
| J | $255^0$ < angle ≤ $285^0$ |
| K | $285^0$ < angle ≤ $315^0$ |
| L | $315^0$ < angle ≤ $345^0$ |

**Algorithm 1**

1. begin
2. Set $RV$ = 0 and $k$ = 1.
3. Repeat steps 3 to 15 until $k \leq N$.
4. Set $j$ = 1.
5. Repeat steps 5 to 14 until $j \leq 19$.
6. Set $d = |D_{I_j} - D_{k_j}| + 1$.
7. If ($d > 7$) then
8. Set $d = |12 - d| + 1$.
9. End if
10. Set $RV_k = RV_k \times |d|$.
11. If ($RV_k < RV$) then
12. Set $RV = RV_k$.
13. Set $RS = D_k$.
14. End if
15. Increment $j$ by 1 and go to step 4.
16. Increment $k$ by 1 and go to step 2.
17. end

In Algorithm 1,

$N$: Total number of strokes in small line segments directions database.

$k$: $k$ is an integer and $1 \leq k \leq N$.

$D_I$: Input handwritten stroke.

$D_k$: $k^{th}$ Stroke of small line segments directions database.

$RS$: Recognized stroke.

$RV$: Recognition value of RS.

$RV_k$: Recognition value of $k^{th}$ stroke, where $1 \leq k \leq N$.

$d, j$: Integers.

## 4. RESULTS AND DISCUSSION

Our focus is to test whether some input handwritten character can be recognized correctly or not in mobiles phone devices. An application in J2ME is developed that implement recognition procedure discussed in Section 3. The evaluation of proposed character recognition system has been performed on Gurmukhi and Devanagiri characters presented in Figure 1 and Figure 2, respectively. The results have been obtained using mobile phone having maximum touch screen size is 52mm as width and 111mm as length, and processor is 434 MHz. The present system takes appx. 0.200 seconds to recognize an input handwritten character. The database has been prepared in view to keep necessary parameters as Author ID, data capturing date and time, character to draw, data with stroke markers and end of character marker. One record of used database in context of Devanagiri script has been presented in Figure 6 for Devanagiri character 'A'. In this figure, a field is separated from other field by symbol ',' and end of record by symbol ';'. The first field value is 'Anuj_1_RH' refers to Author ID and this field include three types of information as

<AuthorFirstName_ID_WritingHand>. The next fields are Date, time, character name and points collected in sequential order of strokes. Each new position of stroke starts with symbol 'S'. As discussed in Section 3, a point is collection of x and y co-ordinates, here each point has referred as X<value>Y<value>. The default co-ordinate system has been used of fourth quadrant as values of Y co-ordinates are zeroes.

## 4.1 Results for Gurmukhi characters

The training set includes 5330 characters and test set include 1640. We have noted that a total number of 1553 words have been recognized correctly and the overall recognition rate has been achieved as 94.69% for online cursive handwriting. The Figure 7 presents a screen shot of application in execution. The present results are acceptable with respect to results observed in literature for devices other than mobile phones and these results cannot be compared as lack of exclusive work done in mobile phone devices from literature point of view.

## 4.2 Results for Devanagiri characters

The results have been obtained using training data sets of 1050 characters and testing data sets of 504 characters. It has been observed that 438 charcters from test set have been recognized correctly. Therefore, the overall recognition rate has been achieved as 86.90%.

There are situations during experimentation stage when mobile device technical shortcomings restrained to achieve desired recognition accuracy. As discussed, we have selected a device at medium range from economic and technical advancements point of view, so that solution achieved in this environment could be accepted easily for high range devices as well.

## 5. CONCLUSION

In this paper, we have recognized online handwritten Gurmukhi and Devanagiri characters in context of mobile phone devices. The application study of mobile phone devices and computers differ in many areas as discussed in Section 1. This study is a step towards development of robust recognition engine for mobile phones where processing speed and memory issues could be managed and handwriting recognition can be done easily. The advancements in hardware of mobile phones will ease the use of handwriting recognition engines where large datasets are required. Gurmukhi and Devanagiri are popular script in India and shares many similarities to other Asian scripts such as Bangla, Gujarati, Telugu etc. We believe that the present study will be helpful in recognition of such similar scripts and use of handwriting recognizers for these scripts in mobile phones.

Anuj_1_RH,13-April-2010,09-46-56,A,SX67.Y-125.X68.Y-123.X70.Y-122.X71.Y-121.X73.Y-119.X74.Y-118.X76.Y-116.X78.Y-115.X81.Y-114.X85.Y-113.X87.Y-113.X90.Y-112.X93.Y-112.X95.Y-112.X98.Y-111.X100.Y-111.X102.Y-111.X104.Y-111.X107.Y-111.X108.Y-113.X109.Y-113.X112.Y-115.X113.Y-117.X115.Y-118.X116.Y-119.X116.Y-121.X117.Y-122.X118.Y-124.X118.Y-126.X118.Y-128.X118.Y-130.X118.Y-132.X118.Y-134.X118.Y-136.X116.Y-140.X114.Y-142.X113.Y-143.X111.Y-146.X109.Y-148.X107.Y-150.X105.Y-152.X103.Y-154.X102.Y-155.X100.Y-156.X98.Y-157.X97.Y-157.X95.Y-158.X94.Y-158.X93.Y-158.X92.Y-158.X91.Y-158.X90.Y-158.X89.Y-158.X87.Y-158.X86.Y-158.X85.Y-158.X84.Y-158.X82.Y-158.X84.Y-159.X89.Y-160.X94.Y-161.X97.Y-161.X100.Y-161.X105.Y-163.X107.Y-164.X109.Y-165.X111.Y-166.X113.Y-167.X113.Y-168.X115.Y-169.X115.Y-170.X115.Y-173.X116.Y-175.X117.Y-177.X117.Y-178.X117.Y-180.X118.Y-182.X119.Y-183.X119.Y-185.X120.Y-187.X121.Y-188.X121.Y-190.X121.Y-192.X121.Y-194.X121.Y-195.X121.Y-196.X121.Y-198.X120.Y-201.X119.Y-201.X118.Y-202.X117.Y-203.X116.Y-204.X115.Y-207.X113.Y-208.X113.Y-209.X111.Y-210.X109.Y-211.X107.Y-212.X106.Y-213.X104.Y-213.X101.Y-213.X99.Y-213.X97.Y-213.X94.Y-213.X91.Y-213.X88.Y-213.X85.Y-213.X82.Y-213.X79.Y-211.X74.Y-208.X71.Y-207.X68.Y-205.X64.Y-203.X63.Y-202.X61.Y-200.X60.Y-198.X59.Y-197.X58.Y-194.X58.Y-193.X57.Y-192.X56.Y-190.X55.Y-189.X55.Y-188.X54.Y-186.X54.Y-185.X53.Y-183.X53.Y-181.X52.Y-180.X51.Y-179.X51.Y1.SX101.Y-157.X108.Y-157.X113.Y-157.X119.Y-157.X125.Y-157.X131.Y-157.X136.Y-157.X143.Y-157.X148.Y-157.X152.Y-157.X155.Y-157.X158.Y-157.X159.Y-157.X160.Y-157.X160.Y1.SX153.Y-113.X152.Y-116.X152.Y-120.X153.Y-123.X153.Y-126.X153.Y-129.X154.Y-135.X154.Y-138.X154.Y-141.X154.Y-144.X154.Y-147.X154.Y-150.X154.Y-152.X154.Y-154.X154.Y-156.X153.Y-158.X153.Y-160.X152.Y-161.X152.Y-162.X151.Y-164.X151.Y-166.X151.Y-168.X151.Y-170.X151.Y-172.X151.Y-174.X151.Y-176.X151.Y-177.X151.Y-178.X151.Y-180.X151.Y-181.X151.Y-182.X151.Y-184.X151.Y-185.X151.Y-186.X151.Y-187.X151.Y-189.X151.Y-190.X151.Y-191.X151.Y-193.X151.Y-195.X151.Y-196.X151.Y-197.X151.Y-199.X151.Y-200.X151.Y-201.X151.Y-202.X151.Y1.SX128.Y-108.X134.Y-108.X139.Y-108.X145.Y-108.X153.Y-108.X159.Y-108.X166.Y-108.X173.Y-108.X180.Y-108.X184.Y-108.X188.Y-108.X189.Y-108.X190.Y-108.X190.Y1.;

**Fig 6: A record from Database for character 'A'**

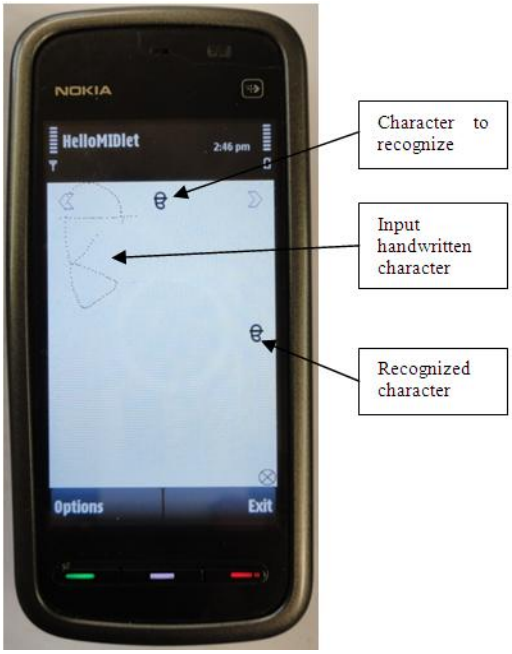


**Fig 7: Recognition system in mobile phone device**

## 6. REFERENCES

[1] Sharma, A., Sharma, R. K. and Kumar, R. 2009. Online Preprocessing of Handwritten Gurmukhi Strokes, International Journal of Machine Graphics and Vision, Vol. 18, no. 1, pp. 105-120.

[2] Sharma, A., Kumar, R. and Sharma, R. K. 2008. Recognizing Online Handwritten Gurmukhi Characters using Elastic Matching. IEEE proceedings of International Congress on Image and Signal Processing (CISP-2008), Sanya, vol. 2, pp. 391-396.

[3] Sharma, A., Kumar, R. and Sharma, R. K. 2009. Rearrangement of Strokes in Recognition of Online Handwritten Gurmukhi Words. In IEEE Proceedings of 10th International Conference on Document Analysis and Recognition, Barcelona, Spain (ICDAR-2009), pp. 1241-1245.

[4] Sharma, D.V and Lehal, G.S. 2006. An Iterative Algorithm for Segmentation of Isolated Handwritten Words in Gurmukhi Script. IEEE proceedings of ICPR.

[5] Sachan, M.K., Lehal, G.S. and Jain, V.K. 2011. A Novel Method to Segment Online Gurmukhi Script. Information Systems for Indian Languages, Communications in Computer and Information Science, Vol. 139, Springer-Verlag, pp. 1-8.

[6] Joshi, N., Sita, G., Ramakrishnan, A. G., Deepu, V., and Madhvanath, S. Machine recognition of online handwritten Devanagari characters. Proceedings of International Conference of Document Analysis and Recognition, pp. 1156-1160.

[7] Bhattacharya, U., Purui, S. K., Shaw, B. and Bhattacharya, K. 2006. Neural combination of ANN and HMM for handwritten Devanagari numeral recognition. Proceedings of ICFHR, pp. 613-618.

[8] Swethalakshmi, H., Jayaraman, A., Chakarvarthy, V. S. and Sekhar. C. C. Online handwritten character recognition of Devanagari and Telugu characters using support vector machines. Proceedings of IWFHR.

[9] Bhattacharya, U., Gupta, B. K., Parui, S., Direction code based features for recognition of online handwritten characters of Bangla. Proceedings of International Conference on Document Analysis and Recognition, vol. 1, pp.58-62.

[10] Babu, V., Prasanth, L., Sharma, R., Rao, G. V., and Bharath, A., HMM-Based online handwriting recognition system for Telugu symbols. Proceedings of the ninth International Conference on Document Analysis and Recognition, vol.1, pp. 63-67.

[11] Prasanth, L., Babu, V., Sharma, R., Rao, G. V., and M., D., Elastic matching of online handwritten tamil and telugu scripts using local features. Proceedings of the ninth International Conference on Document Analysis and Recognition, vol. 2, pp. 1028-1032.

[12] Jain, A.K, Robert, P.W and Mao, J. Statistical Pattern Recognition – A Review. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 1, pp. 4-37.

[13] Sharma, A., Kumar, R. and Sharma, R. K. Recognizing Online Handwritten Gurmukhi Characters using Comparison of Small Line Segments. International Journal of Computer Theory and Engineering, A Journal of World Academy of Computer Science and Information Technology, vol. 1, no. 2, pp. 136-141.

[14] Sharma, A. 2009. Online Handwritten Gurmukhi Character Recognition. PhD Thesis, Thapar University, Patiala.

[15] Beigi, H., Nathan, K., Clary, G. J., and Subhramonia, J. Size normalization in unconstrained online handwritng recognition. Proceedings ICIP, 169-173

[16] Unser, M., Aldroubi, A., Eden, M. 1993. B-Spline signal processing: part II - efficient design and applications. IEEE Transactions on Signal Processing, 41(2), 834-848.

[17] Kavallieratou, E., Fakatakis, N., Kolkkinakis, G. An unconstrained handwriting recognition system. International Journal of Document Analysis and Recognition, 4(4), 226-242.

[18] Brault, J. J. and Plamondon, R. Segmenting handwritten signatures at their perceptually important points. IEEE Transactions on Pattern Analysis and Machine Intelligence, 15(9), 953-957.

[19] Guerfali, W. and Plamondon, R. Normalizing and restoring online handwriting. Pattern Recognition, 26(3), 419.

[20] Bellegarda, E. J., Bellegarda, J. R., Namahoo, D. and Nathan, K. S. A probabilistic framework for online handwriting recognition. Proceedings of IWFHR III, pp. 225-234.

[21] Jain, A.K and Dubes, R.C. Algorithms for Clustering Data. Prentice Hall.