

Handwriting Recognition of Gurmukhi Script: A Survey of Online and Offline Techniques

Ravneet Kaur

Department of Computer Science,
Guru Nanak National College, Doraha, Punjab, India

Abstract: Character recognition is very interesting area of pattern recognition. From past few decades an intensive research on character recognition for Roman, Chinese, and Japanese and Indian scripts have been reported. In this paper, a review of Handwritten Character Recognition work on Indian Script Gurmukhi is being highlighted. Most of the published papers were summarized, various methodologies were analysed and their results are reported.

Keywords: Gurmukhi Character Recognition, Online, Offline, HCR Survey.

I. INTRODUCTION

India is a Multi-Lingual, Multi-Script country. Devnagari, Gurmukhi, Bangla, Tamil, Telgu, Urdu, Oriya, Gujarati, Kannada and Malayalam are 10 official Indian Scripts. Gurmukhi is popular Script used to write Punjabi language. Punjabi language is an Indo-Aryan language spoken by 102 million native speakers worldwide.

Pattern recognition has become interesting research area from last few decades. As compared to handwritten characters, typed characters can be efficiently, accurately and easily recognized by computer machine. Many researchers have done work to recognize machine printed and handwritten characters and many algorithms have been proposed. Researchers have been working on handwritten recognition for more than 30 years. Over the few past years, the numbers of organizations are involved in research on handwritten recognition.

Some practical applications of handwriting recognition are: (i) pen based form filling, (ii) word processing, (iii) natural language processing, reading aid for the blind, (v) language processing, (vi) multi-media system design, (vii) usage of online handwriting recognition in conjunction with speech synthesis, to empower people with vocal disability to communicate with others etc.

II. PROPERTIES OF GURMUKHI SCRIPT

Gurmukhi script is 14th most widely used script in the world. The name Gurmukhi means “from the mouth of the Guru”. Guru Nanak, the first Sikh guru developed the Gurmukhi alphabet during the 16th century. Some properties of Gurmukhi Script are:

- Gurmukhi has 41 consonants (*vianjan*), 9 vowel symbols (*lāga mātrā*), 2 symbols for nasal sounds (*bindī* and *ṭippī*), and one symbol which duplicates the sound of any consonant (*addak*). In addition, four conjuncts are used: three subjoined forms of the consonants Rara, Haha and Vava, and one half-form of Yayya.
- Most of the characters have a horizontal line at the upper part. The characters of words are connected mostly by this line called head line and so there is no vertical inter-character gap in the letters of a word.
- A word in Gurmukhi script can be partitioned into three horizontal zones. The upper zone denotes the region above the head line, where vowels reside, while the middle zone represents the area below the head line where the consonants and some sub-parts of vowels are present. The middle zone is the busiest zone. The lower zone represents the area below middle zone where some vowels and certain half characters lie in the foot of consonants.
- There are many topologically similar shaped characters in Gurmukhi. (ਚ, ਚ, ਚ), (ੲ, ਯ, ਖ), (ੜ, ਝ) are some examples of similar shaped characters.
- The bounding boxes of 2 or more characters in a word may intersect or overlap vertically. For example ਿ and ਚ intersects in word ਿਚ.

All above properties and different writing style of users complicate the Segmentation and recognition of Gurmukhi text.

III. HANDWRITTEN GURMUKHI CHARACTER RECOGNITION

Recognition of handwritten text can be done offline or online. Online handwriting refers to writing through a special pen on an electronic surface. Recognition of Online handwriting is the tracking of hand movements, the way user writes. This type of data is treated as a dynamic representation of handwriting. Offline handwriting recognition (OHR) involves the conversion of handwritten text on an image into a computer readable format. The text in image is considered as a static representation of handwriting. Steps involved in online/offline handwritten character recognition are illustrated in Figure 1: General Steps in Handwritten Character Recognition System.

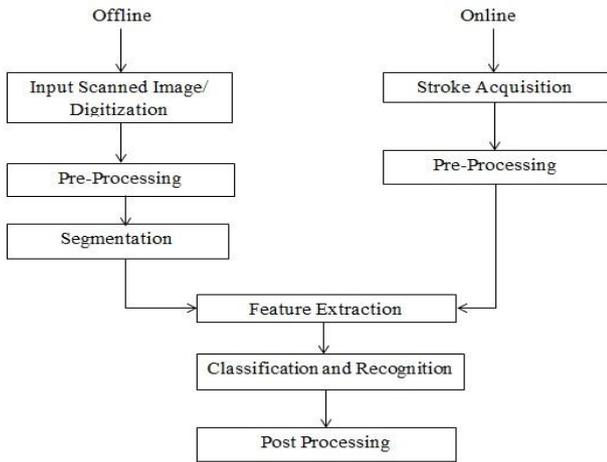


Figure 1: General Steps in Handwritten Character Recognition System

The major difference between Online and Offline Character Recognition is that Online Character Recognition has real time contextual information but in case of offline recognition pattern is stored in the form of image prior to recognition.

IV. WORK FOR GURMUKHI SCRIPT RECOGNITION

Garg (2009) proposed a system for recognition of offline handwritten gurmukhi characters. System is based on structural features, multi-input multi layered Neural Networks are used for the recognition purpose. Accuracy in the range of 69 to 96% is reported [2].

Table 1: Existing Gurmukhi Handwriting Recognition Systems

Author	Input	Features	Method/Classifier	Training Set/ Test Set	Recognition Rate
Naveen Garg (2009)	Offline Handwritten Characters	Structural	Neural Network	-	69% to 96%
Anuj Sharma (2009)	Online Handwritten Characters	<u>High Level Features</u> - Loop, crossing, dots, straight line and headline. <u>Low Level Features</u> - Position of stroke ,area, length, curliness,slope	Small Line Segment Method Based on Elastic Matching and chain code Technique	Test Set 2460 characters from 60 users	94.59% Using Small Line Segmentation
Ubeeka Jain (2010)	Offline Handwritten Characters	Profile, width, height, aspect ratio	neocognitron artificial neural network	Data Set 1500	92.78%
Sharma, Jhaji (2010)	Offline Handwritten Characters	Zone Features	K-NN and SVM	3075/2050 images	72.54%(K-NN), 72.83%(SVM RFB Kernel)
Manoj K Sachan (2011)	Online Handwritten Words	Distributed Directional Features	(a)Extraction of strokes (b) Merging of Sub strokes	Test Set 2150 words from 50 users	76%
Shivali (2011)	Online Handwritten Numerals	Points Generated after pre-processing phase is used as feature for recognition	Support Vector Machine	80% of Data/ 20% of Data	97.5% (Using linear kernel)

A Sharma (2009) introduced a writer independent system for online Gurmukhi script recognition. A new recognition method, Small Line Segment was introduced which was based on chain code rule and elastic matching technique. Recognition rate of 94.59% was achieved using small line segments method. Post processing phase is also discussed which includes verification of recognized strokes through features of character in the script [3].

D Sharma(2010) developed a System for Recognition of Hand-filled forms in Gurmukhi Script. Methodology consist of template generation, skew correction, data extraction, segmentation , slant correction, size normalization, feature extraction, classification and post processing steps[4].

Sharma and Ubeeka Jain (2010) presented a system for recognition of isolated handwritten Gurmukhi characters. Horizontal and vertical profiles, stored height and width of each character and neocognitron artificial neural network are used for feature extraction and classification. Average accuracy of 92.78% is reported [5].

System for Recognition of offline Isolated Handwritten Characters in Gurmukhi Script is proposed by Sharma and Jhaji (2010). Feature extraction method Zoning is used for extracting features. Recognition accuracy of 72.54% using KNN and 72.83% using SVM (RBF Kernel) is reported [6].

K S Siddharth (2011)	Offline Handwritten Characters	16 zoning density features, 128 background directional distribution features	SVM with RBF Kernel	7000 Samples	95.04%
Munish Kumar (2011)	Offline Handwritten Characters	Diagonal, intersection and open end point features	Support Vector Machine	3500 images 90%/10%	94.29%
Pritpal S (2012)	Offline Handwritten Characters	Feature Extracted using different wavelets	Back Propagation Neural Network(BPNN)	400/100 Samples	94.41%
Gita (2012)	Offline Handwritten Characters	Zone Based Feature Extraction	K-NN and SVM	Data set 7000 samples	95.11%(SVM) 90.64%(K-NN)
S Singh (2012)	Offline handwritten Characters	Gabor Features-GABM Gabor Features-GABN	SVM with RBF Kernel	200 Samples	94.29%
Aggarwal (2012)	Online Handwritten Strokes	Points Generated after pre-processing phase is used as feature for recognition	Support Vector Machine	Test Set 300 words from 3 users	98.45%(without pre-processing) 98.92%(with Pre-processing)
Khurana (2013)	Online Handwritten Characters	Points Generated after pre-processing phase is used as feature for recognition	Support Vector Machine	75% of Data/ 25% of Data	94.4%

Sachan, Lehal and Jain (2011) proposed a system for online Gurmukhi script recognition. The segmentation algorithm consist of two phases namely extraction of strokes and merging of strokes. Segmented shapes are processed to extract Distributed directional features. Feature data is fed to the recognition engine which is nearest neighbour classifier. Average accuracy of 76% is achieved by the system [7].

Shivali (2011) has reported work for recognition of Online Handwritten Punjabi Numerals using Support Vector Machine. Recognition accuracy of 97.5% is achieved with 40 resampled points using linear kernel [8].

System for offline Gurmukhi character recognition using zoning density and background directional distribution features is presented by Siddharth, Dhir and Rani (2011). SVM classifier with RBF kernel is used for classification. Average accuracy of 95.04 is reported [9].

A scheme for offline handwritten Gurmukhi character recognition based on SVMs is presented by Munish, Jindal and Sharma (2011). Maximum recognition accuracy of 94.29% with 90% training data and 10% testing data using intersection and open end points as features is reported [10].

Pritpal S and S Budhiraja(2012). Method uses different Daubechies Discrete Wavelet Transforms for feature extraction and back propagation for recognition task. Average accuracy of 93.41% is reported [11].

Gita, Rani and Dhir (2012) presented a system using K-NN and SVM classifier for offline handwritten gurmukhi character recognition. Recognition accuracy of 95.11% using SVM and 90.64% using K-NN is reported [12].

Singh, Aggarwal and Dhir (2012) used Gabor filters for recognition of offline handwritten

Gurmukhi characters. Recognition accuracy of 88.27% using GABM and 94.29% using GABN is reported [13].

Aggarwal (2012) has been reported work for recognition of Online Handwritten Gurmukhi Strokes using SVM. An overall accuracy of 98.45% without applying pre-processing and 98.92% with pre-processing phase has been reported [14]. Khurana (2013) has also reported work for online Gurmukhi text recognition by using SVM with accuracy rate of 94.4% [15].

Little work has also been reported for Handwritten Character Recognition for Mobile Devices. A. Sharma (2012) proposed a system to recognize online handwritten Gurmukhi and Devanagari characters in touch screen based mobile phones. They have used small line segments to recognize Gurmukhi and Devanagari characters [16].

Table 1: Existing Gurmukhi Handwriting Recognition Systems gives an overview of the online and offline handwriting recognition systems for Gurmukhi characters and numerals. Researchers tried to explore the possibilities of directional, structural, zonal, statistical features. The majority of the researchers have used neural network, K-NN and SVM based classifiers. There is lack of availability of benchmarking databases for research. From the literature, it is clear that many of the researchers have used their own databases for evaluating their techniques. So it is very difficult to compare the various techniques and methods proposed by the researchers.

V. CHALLENGES

There are many challenges for identifying handwritten scripts. Script identification from printed documents is much easier than handwritten documents. Handwritten script identification is

complex and difficult because the writing style varies from person to person. It is also affected by the state of mind, mood of the person, writing medium, environment etc. Because of this, it is a challenging task to find proper features for their accurate identification. Presence of some similar shape characters (e.g. ਚ, ਦ, ਢ) creates other challenges in the identification. Due to limitations of the device or speed of writing, it is possible for a single character to be broken into different parts, hence creating confusion in recognition. In case of offline handwriting recognition there is a problem of degraded documents, poorly scanned documents make the recognition task challenging.

VI. FUTURE WORK

The work reported on HCR for Gurmukhi script may be extended in various directions. Some of them are listed below:

- Little work has been reported toward the recognition of handwritten Gurmukhi words.
- Development of bilingual and multilingual HCR.
- Incremental learning can be incorporated so that a new stroke or character can be incorporated without the requirement of retraining the entire system.
- Extension of handwriting recognition to the recognition of large or very large vocabularies.

VII. CONCLUSION

This paper summarizes work on Gurmukhi handwritten text recognition. Many of the researchers do research for a short period of time as a part of their academic courses. So there is a great amount of redundancy in the work related OHR. Although research in the field of HCR has been done, recognition algorithms still do not achieve 100% of the high recognitions.

REFERENCES

1. G.S Lehal, C. Singh, "A Gurmukhi Script Recognition System", In Proc IEEE, pp. 557-560, 2000.
2. N. Garg, "Handwritten Gurmukhi Numeral Recognition using Neural Network", M.tech Thesis, Thapar University, Patiala.
3. A. Sharma, "Online Handwritten Gurmukhi Character Recognition", Doctor of Philosophy Thesis, School of Mathematics and Computer Applications, Thapar University, Patiala, 2009.
4. D. Sharma, "A System for Recognition of Hand-filled forms in Gurmukhi Script", Doctor of Philosophy Thesis, Punjabi University, Patiala 2010.
5. U. Jain, D. Sharma, "Recognition of Isolated Handwritten Characters of Gurmukhi Script using Neocognitron", International Journal of Computer Applications (0975-8887), Vol. 4, No. 8, 2010.
6. P. Jhaji, D. Sharma, "Recognition of Isolated Handwritten Characters in Gurmukhi Script", International Journal of Computer Applications (0975-8887), Vol. 4, No. 8, pp. 09-17, 2010.
7. M.K Sachan, G.S Lehal, V.K Jain, "A System for Online Gurmukhi Script Recognition", ICISIL, pp. 299-300, 2011.
8. Shivali, "Recognition of Online Handwritten Punjabi Numerals", Master of Science Thesis, School of Mathematics and Computer Applications, Thapar University, Patiala, 2011.
9. K.S Siddharth, R. Dhir, R Rani, "Handwritten Gurmukhi Character Recognition Using Zoning Density and Background Directional Distribution Features", International Journal of Computer Science and Information Technologies (0975-9646), Vol. 2 (3), pp.1036-1041, 2011.
10. M. Kumar, M.K Jindal, R.K Sharma, "SVM Based Offline Handwritten Gurmukhi Character Recognition", International Conference on Image Information Processing (ICIIP 2011).
11. P. Singh, S. Udhiraja, "OCR for Handwritten Gurmukhi Script using Daubechies Wavelet Transforms, "International Journal of Computer Applications (0975 – 8887) vol. 45- no.10, pp. 25-28, 2012.
12. G. Sinha, R. Rani, R. Dhir, "Handwritten Gurmukhi Character Recognition Using K-NN and SVM Classifier", International Journal of Advanced Research in Computer Science and Software Engineering. vol. 2, issue 6, pp.288-293, 2012.
13. S. Singh, A. Aggarwal, R. Dhir, "Use of Gabor Filters for Recognition of Handwritten Gurmukhi Character", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 2, Issue 5, pp. 235-240, 2012.
14. R. Agrawal, "Recognition of Online Handwritten Gurmukhi Stroke using Support Vector Machine", M.tech. Thesis, School of Mathematics and Computer Applications, Thapar University, Patiala, 2012.
15. D. Khurana, "Online Handwritten Gurmukhi Character Recognition using Support Vector Machine", M.tech. Thesis, School of Mathematics and Computer Applications, Thapar University, Patiala, 2013.
16. A. Sharma, K. Dahiya, "Online Handwriting Recognition of Gurmukhi and Devanagiri Characters in Mobile Phone Devices", International Conference on Recent Advances and Future Trends in Information Technology (iRAFIT2012), pp. 37-41, 2012.
17. U. Pal, B.B Chaudhuri, "Indian Script Character Recognition: a survey", Pattern Recognition 37, pp. 1887–1899, 2004.
18. U. Pal, R. Jayadevan, N Sharma, "Handwriting Recognition in Indian Regional Scripts: A Survey of Offline Techniques". ACM Transactions on Asian Language Information Processing, 2012.