A Comprehensive Review on Handwriting Character Recognition

Dr. S. Vijayarani

Assistant Professor, Department of Computer Science,
Bharathiar University, Coimbatore, India
Vijimohan_2000@yahoo.com

A. Sakila

Research Scholar, Department of Computer Science,
Bharathiar University, Coimbatore, India
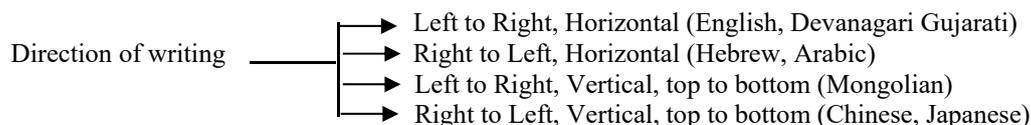Sakivani27@gmail.com

A. Revathi

PG student, Department of Computer Science,
Bhrathiar University, Coimbatore, India
revathi1621998@gmail.com

Abstract---Nowadays handwritten character recognition plays a major role and becomes one of the challenging research domains in document image analysis. Normally, every individual's handwriting is unique which is proved by the recent studies and handwritten characters are hard to recognize because of various different handwritten styles, sizes, shapes and angles followed by humans. A present, handwritten/signature forgery has become big and common issue. Hence, recognition and authenticity verification of the handwritten characters of particular individual is a challenging issue. The way of holding pen and amount of pressure put on paper differ from everyone that cannot be mimicked or forged. Currently, Handwritten Character Recognition domain is useful for plenty of applications. During the early days, it has been used for data entry, application for loans, bank cheque, authentication of signatures in banks, credit cards, tax, insurance of health forms, passport validation and so on. This paper aims to present the comprehensive review on basics of handwritten character recognition, offline and online handwritten character recognition system, various process involved, applications, issues and challenges in handwritten character recognition system.

Keywords-Handwritten Images, Handwritten Recognition, Online Handwritten Recognition, Offline Handwritten Recognition.

I. INTRODUCTION

Handwritten Character Recognition (HCR) is a challenging research area in document image analysis, which defines an ability of a machine to mine the patterns and identifies character/word/sentence from various sources such as paper photographs, documents, touch screens and similar devices by a computer. Nowadays HCR is used in communication and transmission media i.e. word processors, fax machines and email. Handwritten differs from writer to writer because everyone has their own writing style. Sometimes it is very difficult for human being to identify and read the handwritten texts of another human.HCR can be associated with more types of recognition and interpretation. Identification of the handwritten images is used to determine the sample handwriting from set of writers. HCR technology is divided into two different categories for recognize the word/character/sentence from handwritten images, they are online and offline. In offline handwriting recognition is used to find the letters that are present in the digital image of handwritten text or it can be from a piece of a paper by optical scanning or intelligent word recognition. Online techniques detect the motion of the pen tip, a pen based computer screen surface that is data are captured during the writing process with the help of a special pen. The recognition rates are higher for online case than the offline case. The online case makes attractive, so it is considered for development whereas the offline case first estimates the path of writing from offline data and then uses the online algorithms.

Direction of writing
- Left to Right, Horizontal (English, Devanagari Gujarati)
- Right to Left, Horizontal (Hebrew, Arabic)
- Left to Right, Vertical, top to bottom (Mongolian)
- Right to Left, Vertical, top to bottom (Chinese, Japanese)

Figure 1. Left to right (horizontal)

Figure 2. Right to Left (horizontal)

Figure 3. Left to Right (Vertical)

Figure 4. Right to Left (Vertical)

The remaining portion of this paper is explained as follows. Online and offline handwritten character recognition are given in Section 2. Section 3 discusses the handwritten character recognition architecture. Applications are presented in Section 4. Research issues and challenges are given in Section 5. Section 6 shows the conclusion.

II. ONLINE AND OFFLINE HANDWRITTEN CHARACTER RECOGNITION

Handwriting Character Recognition (HCR) technology is an inventive challenging domain, from the handwritten images the HCR identifies and recognizes the text. HCR is classified into two different techniques; i.e. Online and Offline handwritten character recognition system.
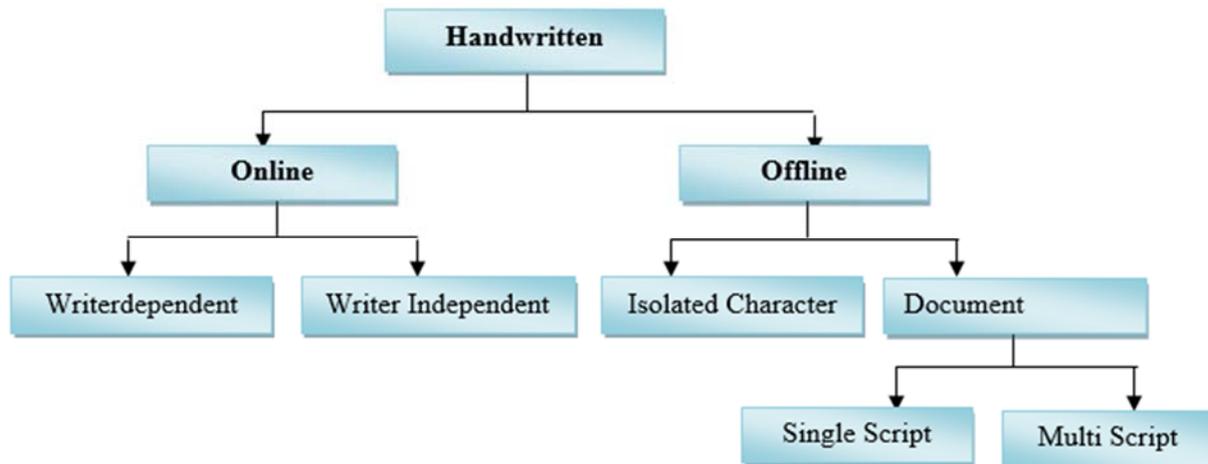

Figure 5: Online and Offline Handwritten Character Recognition

A. Online Handwritten Character Recognition system

In this online handwritten character recognition technique, the directional information means to identify in which direction the user wrote (left to right, right to left, top to bottom and bottom to top. From this we can get the basic information about character, unique style of writing, use of the pointing objects.Pointing object may go from left to right, right to left, top to bottom and bottom to top directional. Main function of online handwritten character recognition system is the classification process. It classifies the handwritten character and stored as ASCII code format for further processing. There are two types of online handwritten character recognition system; (i) writer dependent and (ii) writer independent.

1) *Writer Dependent*: In Writer Dependent online handwritten character recognition technique, first the end user provides the basis structures of input to recognition engine.
2) *Writer Independent*: In Writer Independent online handwriting character recognition technique, the user can start using system without worry about underlying recognition engine.

B. Offline Handwritten Character Recognition System

Offline handwritten character recognition technique mine the handwritten paper and it can be stored as in the form of image. Consider the scenario where one user send a handwritten scanned digitized copy through mail to another person if the end user who have received it want to do modification in that document for further processing but it cannot be done because image doesn't allow text searching, editing and indexing. The task of identifying handwritten character in image is converted into specific format which allows reformat for further processing. It does not really require user to have operating knowledge of device and it require into online handwritten character recognition system. It can be further divided into Isolated Character Recognition and Document Recognition on the basis of how much content is stored in it.

1) *Isolated Character Recognition:* In this category image contains single character written in any script or symbol which is little easy when compared to document recognition. The system does not require segmentation of sentences as character and word is in isolation form.
2) *Document Recognition:* Document may have multiple pages which may have multiple lines, further divided into words and characters. Offline handwritten document recognition engine requires segmentation to break apart pages to lines, lines to words, and words to character then it focus the feature extraction and classification process. There are two typesof document recognition, they are, Single script and Multi script on the basis of language script.
a) *Single Script:* The document is written with single language script then it is very easy to identify and classify the features then it convert into ASCII equivalent.
b) *Multi Script:* The document is written with multiple script or mixture of multiple language script then recognition engine have to deal with feature extraction and classification with all alphabets used for the document which makes challenging than single script recognition.

III. HANDWRITTEN CHARACTER RECOGNITION ARCHITECTURE

Handwritten Character Recognition architecture is given in Figure 1. The architecture has several components. The important components are Input handwritten Images, Optical Scanning, Preprocessing, Segmentation, Feature Extraction, Classification, Linguistic Processing and ASCII Sentences.
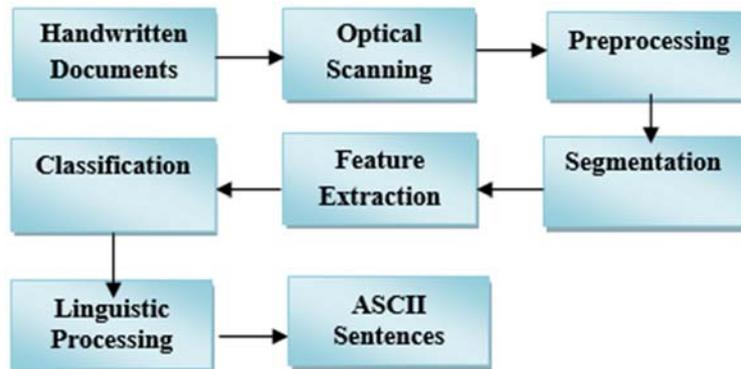

Figure 6. Handwritten Character Recognition Architecture

A. Input Handwritten Documents

Input handwritten documents should be in paper format which is obtained from various sources like student's answer sheet from school and college and also other paper formats like postal letter, bank challans/cheque, and government records and so on.

B. Optical Scanning

In this case of optical scanning the input is obtained by analyzing the numerals of thecheque which is placed at particular location on the scanner. The image of the numerals and its immediate neighbor is converted into binary bitmap values, "-1" means "on" and "0" means "off". Mostly the cheque contains box or line where the amount should be written in numerical format.

C. Pre-processing

The important task is pre-processing, which is carried out by scanned input image. The scanned documents may require some preliminary processing. Pre-processing helps to produce the final document which will be processed by handwritten character recognition system. The major operations involved in pre-processing are Normalization, Binarization, Skew detection and correction.

1) Noise Reduction (Normalization)

Noise Reduction or normalization is performed in the document analysis task which deals with the typed documents or machine printed documents. For the handwritten the connectivity of strokes has to be preserved. Noise removal is the process of removing the noise from the given image. Noise reduction application used to increases the quality of the document. Filtering and Morphological operations are used to perform the noise removal. In digitization, various types of noise are occurs in handwritten images. There are salt and pepper noise, Gaussian noise, Gamma noise, Uniform noise. After normalization, the noise can be removed from the input images. For example, without losing the data, thinning shape of a character can be gathered.

2) Binarization

In this binarization the gray scale image is transformed to a binary image that is black pixel as foreground and white pixel as background using suitable threshold. It can be further divided into global and adaptive binarization. Global Binarization uses the threshold value for the entire document image, the value is picked up and the picking value is based on the estimated value of the image. In Adaptive Binarization each pixel uses different values.

3) Skew detection and correction

Skew detection and correction is used to correct the text line in scanned document images which is not horizontally aligned during the scanning process. Slant is one of the skew type which is differ from everyone because every person's slant of handwritten is different hence it is difficult to identify. The normalization technique uses the skew detection to correct the slant of the writing style.

D. Segmentation

The noise free image from the pre-processing is send to segmentation process. There are three segmentation techniques are used in handwritten recognition, they are, line segmentation, word segmentation, and character segmentation. There is variation among the handwritten so each line of the handwritten images may not be perfectly horizontal that's why the line segmentation is used to align the lines horizontally. The word segmentation

is used to separate the text region into the unit of lines and then finally into words and then words is split into characters using character segmentation.

E. Feature Extraction

Feature extraction plays major role in handwritten recognition. It can be described to extract the representative information from handwritten images. Feature extraction techniques are used to extract features from the handwritten image. Here holistic and analytic are the two main approaches of feature extraction. In holistic recognition, every word is considered to be a class and it is recognized as whole word. The analytic approach is based on character segmentation- free recognition. In Feature selection the most required features improves the classification accuracy and this task is performed by removing irrelevant or noisy features from the entire set of the available ones. The feature is further classified into two different types as Statistical feature and Structural feature.

- *Statistical features* works are based on the mathematical representation. The statistical features are used to represent the handwritten image and it takes the style variations to some extent. Mainly this type is used to reduce the dimension of the feature by setting high to low speed complexity but it will not allow the reconstruction of the original image.
- *Structural features* works are based on various local and global properties of character which is represented by topological and geometrical properties with style variations and high tolerance of distortions.

F. Classification

Classification is used to identify which category the new observations belongs and it is the decision making part. On the basis of training data, the data contains observations. Most commonly seen classifiers are K-Nearest Neighbor (K-NN), Neural Network (NN) and SVM. The classifiers compare the given vector with the stored pattern and give the excellent match as a result.

G. Linguistic Processing

Once the classification process is completed, the next step is to refine the result. The linguistic processing is used to increase the recognition rate, and it is used to detect and correct grammatical misspellings output result.

H. ASCII Sentences

The result of the input handwritten images is displayed as the output. This is stored as ASCII format for making in editable form.

IV. APPLICATION AREAS OF HCR

Nowadays HCR domain is useful in plenty of application.During the early days, it has been used for data entry, application for loans, bank cheque, authentication of signatures in banks, credit cards, tax, insurance of health forms, passport validation and so on.

A. Banking

Nowadays technology is rapidly increased but still bank cheque plays a big role in non cash transaction even after arriving of credit cards, debit cards and electronic transactions. Still in many developing countries, bank cheques are processed manually on day by day. So, the Image processing techniquesand Document image analysis techniques are used, to checkthe handwritten bank cheque.

B. Intelligent Character Recognition

Intelligent Character Recognition is the latest application in image processing it is nothing but the computer translation is used to convert the handwritten text into machine readable and editable characters.

C. Signature

In personal verification in financial circle the handwritten signature plays a big role. The demand of authorization based on signature is increased including credit card validation, security systems, banking system, cheques and so on. Application of HCR is used to reduce fraud in all related financial transaction sectors.

D. Health care

The uses of HCR in health care is to increase the efficiency of patient health details. Although the health care professionals have to deal with large volumes of records for anayzing the patient details including insurance as well as general health forms.As a result, healthcare providers focus on delivering the best possible service to every patient.

E. Passport validation

Since, the number of theft occur in passport verification. Each passport generates the unique MRZ(Machine Readable Zone)code based on the input given by the individual and it matches with the client. The generation of MRZ validated with passport verificationlike passport details which are given by the user while applying passport

and also user provide an image of passport it feed into system to get the details using optical character recognition technology.

V. CHALLENGES AND ISSUES IN HANDWRITING CHARACTER RECOGNITION

Identifying and extracting of handwritten character is still playsa major role,because of vaious handwritten technique. There are still difficulties while reading cursive and non-constrained handwriting, for example some persons writes "1" as "7" and others writes "7" as "4". Some people doesn't care about handwriting style while filling in forms, and most of the time people will write out of the textfields on document filling process than the within text fields. Some persons can write contonuously without any separation. So, it is also difficult to read. Most challenging problem is processing of poor quality handwriting. Cursive words can also significantly reduce the the processing speeds because it is little bit difficult to understand.

VI. CONCLUSION

Handwriting Character Recognition (HCR) is a major research domain in the field of image processing, which is used to identify and recognize the text from handwritten images. Handwritten character differs from every individual and it has the certain level of individuality and uniqueness like DNA. The way of holding pen and amount of pressure put on paper is differ from every individual that cannot be mimicked of forged. This paper elaborately described the handwritten character recognition basic concepts, types, architecture, application area, challenges and issues.

REFERENCES

[1] Plamondon Rejean, and Sargur N. Srihari. "Online and offline handwriting recognition survey. "Pattern Analysis and Machine Intelligence, IEEE Transactions.
[2] Park, Jaehwa, Venu Govindaraju, and Sargur N. Srihari. "OCR in a hierarchical feature space. "Pattern Analysis and Machine Intelligence, IEEE Transactions.
[3] Srihari, Sargur N., et al. "Individuality of handwriting, "Journal of Forensic Sciences 47.4 (2002): 856-872.
[4] Tappert, Charles C., Chilg Y. Suen and Toru Wakahara, "The state of the art in online handwriting recognition. "IEEE Transactions on pattern analysis and machine intelligence 12.8 (1990): 787-808.
[5] P. Shankar Rao, J. Aditya {Dept of CSE, Andhra University}. "Handwriting Recognition – "Offline" Approach.
[6] Manoj Sonkusare and Narendra sahu Dept on CS & Engg., Ujjain Engineering college, ujjiain india and Dept of CS & Engg., Women's Polutechnic college, Indore, India "A survey on handwritten recognition (HCR) Techniques for English alphabets.
[7] Arica, afiz and Fatos T. Yarman-vural. "An overview of character recognitionfocused on offline handwriting systems"
[8] Csey, Richard G., and Eric Lecolinet, "A Survey of methods and strategies in handwriting recognition" Pattern Analysis and machine Intelligence.
[9] Anil. K. Jain Torfinn Taxt, "Feature extraction methods for character recognition – A Survey pattern recognition, vol. 29.
[10] Choudhary, Amit, Rahul Rishi and savita Ahlawat, "Offline Handwritten character recognition using feature extracted from binarization techniques. "AASRI procedia 4 (2013): 306-312
[11] Vijay Prasad, Y. Jayanta singh, Dept of CS and IT, Don Bosco college of engineering and technology, Assam Don Bosco University, Airport road, Azara, Guwahati, Assam "A study on structural method of feature extraction for handwritten character recognition.
[12] Oivind Due Trier, Anil K Jain, Torfinn taxt, "Feature extraction methods for character recognition – A Survey.
[13] Pratibha A. Desai, Sumangala N. Bhavikatti" and Rajeshekar Patil", Dept. of Electronics & communication Engineering, SDM college of engineering and technology, Dharwad, India, "Design and simulation of handwritten text recognition system.
[14] Hsin-chia Fu, Member, IEEE, and Yeong Yuh Xu "Multilinguistic handwritten character recognition"
[15] Elie Krevat, Elliot Cuzzillo, "Improving offline handwritten character recognition with hidden markov models, "2005".