

A Systematic Review of Automation in Handwritten Character Recognition

Madhuri Maheshwari

*Research Scholar, Department of Electronics and Communication
Gurukul Institute of Engineering and Technology Kota, Rajasthan, India.*

Deepesh Namdev

*Faculty, Department of Electronics and Communication
Gurukul Institute of Engineering and Technology Kota, Rajasthan, India.*

Saurabh Maheshwari

*Faculty, Department of Computer Engineering
Govt. Women Engineering College Ajmer, Rajasthan, India.*

Abstract:

Handwritten character recognition has been a problem of interest for the researchers for the last few decades still the problem needs practically implementable and scalable solutions to different applications. One of the main challenges in this field is to recognize the cursive handwritten text. The problem is to align the text, segment individual characters and then recognize the characters. In this paper, we describe the needs, applications, basic terminology and specific research contributions related to handwritten character recognition. We line-out steps involved in the process and also classify various methods related to each step. We also provide advantages and disadvantages of various feature selection and classification methods. We conclude with the analytical comparison of the methods and provide the reader with a systematic review of this area. We also mention the future directions towards an optimal handwritten character recognition method.

Keywords: Handwritten character recognition, handwriting recognition, optical character recognition, hidden Markov model, segmentation, character recognition, classification.

INTRODUCTION

Handwriting is a skill that is unique for each person [1]. Handwriting was developed to facilitate communication and to expand human memory. Writing is made possible much of civilization and culture. Each script contains a set of icons. These icons are known as characters or letters. Each letter has a specified shape. Handwriting has three fundamental characteristic such as:

- Handwriting consist artificial graphical marks on surface.
- Purpose of handwriting is to communicate something
- This purpose is achieved by purity of mark's conventional relation with language.

Many types of analysis, recognition, identification and interpretation associated with handwriting. Central task of handwriting recognition involve word recognition and character recognition. Handwriting recognition provides automatic solution for large amount of data, e.g. postal code reading, mail sorting, many scientific and business applications. Other important uses of handwriting recognition are annotating, editing and other applications that use direct pointing and manipulation [2]. The task of handwriting reading involves specialized human skill. Handwriting interpretation and recognition are used to filter out the variation of messages.

- **Handwriting Recognition:** It is a method of transforming a language which is represented in its spatial form of graphical marks into its symbolic representation [2].
- **Handwriting Interpretation:** It is the process of determining the meaning of handwriting. For example- handwritten address interpretation [3].
- **Handwriting Identification:** By assuming that each person's handwriting is unique, determine author of the handwritten document from the set of authors. For e.g. signature verification which is the method of determining, whether or not the signature is that of a given person [3].

Schemes of Optical Character Detection:

Character detection also called as optical character detection/recognition (OCR) because it performs operation with optically processed characters. Major schemes that exist in OCR are as follows:

- **Detection of Fixed Font Character** is the specific font identification (OCR-B, OCR-A, Elite, Pica, and Courier etc.) type written characters [4].
- **Detection of Multi Font Character** recognizes greater than one font [5].
- **Detection of Omni Font Character** used in identification of any type of font [5].

- **Detection of On-Line Character** takes identification of handwritten single character in which each stroke of character image is provided with timing information [4].
- **Detection of Handwritten Character** is the identification of handwritten single character. That is not written in good handwriting and unconnected [4].
- **Script Detection** is identification of handwritten connected and cursive character which is unrestricted [5].

Techniques of Character Recognition:

The methods used in character identification are same as image analysis and pattern recognition of general image pattern recognition problem [4]. For character identification problem techniques can be summarized as follows:

- **Global Comparison Point by Point:** In this technique point by point all the pixels are compared which are stored in image.

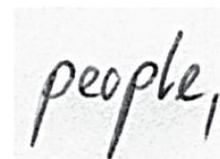
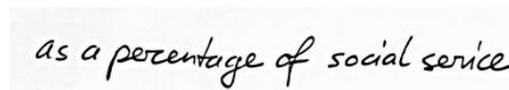
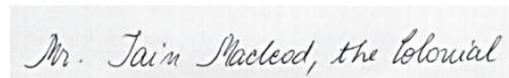
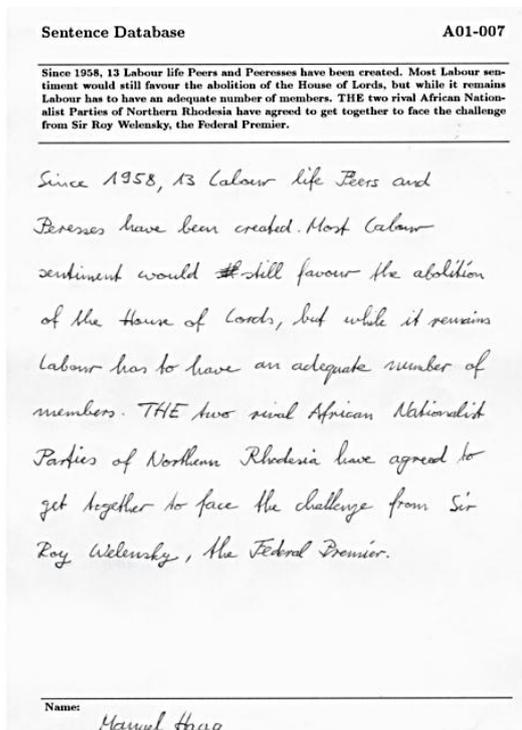


Figure 1: Sample of Handwritten Script of IAM Database [6]

- **Global Transformation:** It takes Fourier, moment calculation then finally rotates with the major axis of inertia.
- **Local Properties Extraction:** Local characteristic such as corners or angles, endpoints, crossing of strokes and T-junctions. Various methods used to extract these properties, for example- initially thin the character and apply masks of variable positions or fixed masks.
- **Extraction of Certain Lines by searches:** Extracted lines can be in the form of horizontal or vertical or rectangles.
- **Means of Curvatures Analysis:** It includes detection of geometrical analysis and concavities.
- **Structural Methods:** In this method calculate constituent element of character by decomposition then use topological reduction and description of the character in the form of a graph.

Datasets: Dataset is a collection of large number of handwritten script which is written by different-different writers. This script is used to generate ground truth for

handwriting recognition by creating trained network. Databases of many languages are available such as Roman/Latin, Chinese, Japanese, East Asian, Arabic and Indian languages. Some Roman databases are IAM, RIMES, NIST, MNIST, CEDAR, IRONOFF, RODIGRO, IBM UB, CVL, firemaker and some Arabic databases are IFN/ENIT, ADAB, AHDB, ARABASE and CENPARMI. Fig. 1 shows the sample of handwritten script of IAM database [6].

HANDWRITING RECOGNITION STEPS

Handwriting recognition uses a top-down method. Fig. 2 shows the steps used in handwriting recognition. It starts with handwritten input by offline or online. To clear the image use some preprocessing methods then segment document into lines, word and sub-words. For handwriting recognition extract feature from segmented word which are used in testing. Last and final step is post-processing that is syntax, semantic and lexical correction.

1. **Handwritten Document:** It is the collection of alphabet of letters / characters in the form of words or

sub-words. Essential property of writing that makes communication feasible is that different characters have significant difference between them than dissimilarity between drawings of similar characters. Handwritten document can be categorized into three categories on the basis of drawing of characters in words.

- **Discrete Character:** This type of character can be differentiated by difference between characters of each word that is boxed discrete character and spaced discrete character [7]. Fig. 3 shows the example of boxed and spaced discrete characters.

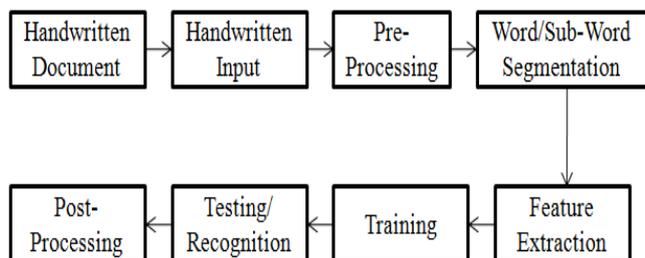


Figure 2: Block Diagram of Handwriting Recognition

- Boxed Discrete Character:** In this type of characters separation of character provided by writers.
- Spaced Discrete Character:** These characters contain one or more stroke between each character.



Spaced Discrete Characters

Figure 3: Boxed and Spaced Discrete Characters [7]

- **Cursive:** In cursive script several characters made within one stroke [7] as fig. 4.



Figure 4: Cursive Script [7]

- **Touched:** Fig. 5 shows the touching character which involve two characters or more than two characters which are uncommon.[8]

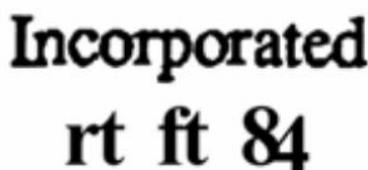


Figure 5: Touched Character [8]

- Handwritten Input:** Handwritten script is changed into digital view either by using unique pen for writing on electronic plane or by scanning the writing paper. These two techniques differentiated by on-line and offline.

- **On-line:** It pertains availability of trajectory data during writing means the 2-D coordinates of consecutive points of writing are stored in order as a time function. Strokes order is readily available and deals with handwritten input spatio-temporal illustration [2]. A transducer used in on-line handwriting detection to capture the writing when it is written. Devices used in on-line handwriting recognition are digitizer or electronic tablet that has 200points/in resolution, 100points/s sampling rate and pen down or inking indication [7].

Digitizer Technology:

Tablet accepts both writing and graphics because it is a powerful mechanism for drawing and sketches input. Digitizer tablet used for many tasks such as position selection, orientation, quantify, path and text input. Two main technologies are used by tablet digitizers that are pressure sensitive and electrostatic or electromagnetic.

- Electrostatic or Electromagnetic Tablets:** It contains 0.1 - 0.5 space in a wire loop and tablet in the stylus tip with x, y conductors grid. Detects the produced current or voltage in sinusoidal signal by acceleration of an electrostatic pulse in grid of conductor or wire loop.
- Pressure-Sensitive Tablets:** It consists a mechanical spaced layer of resistive and conductive material. In x or y route an electrical potential is applied on one of the resistive layer which sets a voltage corresponding to the position. Conductive layer picks the voltage from resistive layer which occurs by stylus tip pressure at a point [7].

- **Off-line:** Scanned images of completed handwriting available in off-line handwriting detection. It deals with spatio-luminance representation of handwritten input [2]. Optical scanners are used in handwritten document scanning such as divided slit scan, mechanical scanner, laser beam scanner, flying spot scanner, vidicon scanner and photocell matrix scanner [9]. An optical scanner converts the image of handwritten script into a bit pattern. Scanners contain x and y resolutions of typically 300-400 points/in [7].

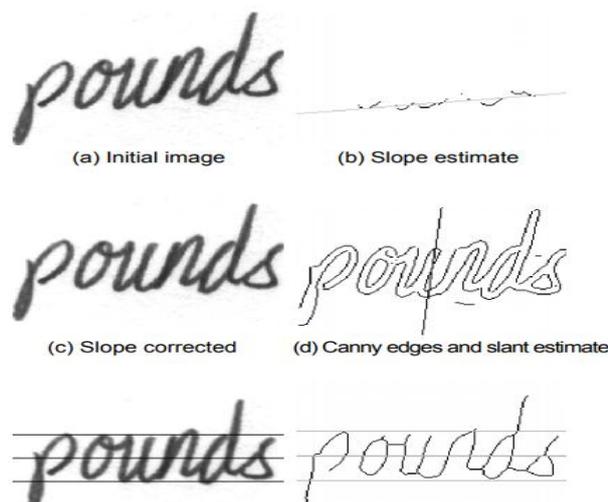


Figure 6: Several Steps of Pre-Processing [10]

HANDWRITING RECOGNITION TECHNIQUES

Language is process of communication. There are four processes of language which is listening, reading, talking and writing. A unique mode of learning is represented by Writing. Writing originates and creates a distinctive unwritten construct that recorded graphically. We learn by 3 major ways that is “by doing”, iconic and represent in words [1]. Handwriting is the way of writing with hand using writing instrument such as pen or pencil. A language contains an alphabet of letters or characters. Writing makes difference between different characters and characters which have same drawing. Handwriting is the stroke’s time sequence. Stroke occurs by pen-up to pen down movement during writing. Each person has a unique way of writing. Handwriting recognition (HWR) is the transforming method of a language into symbolic representation from spatial form of graphical symbols. Handwriting identification includes word recognition or character recognition [2].

HWR includes eight steps which are database collection, image acquisition, skeleton detection, sub-word detection, feature extraction and training of network, testing/recognition and post-processing. A database is collection of handwritten script and words. Some databases are IAM (collection of full English sentences) [12], CVL (1 German and 6 English text) [13], IFN/ENT (Tunisian town and village names) [14], ICDAR, RIMES [15] and PHOND. In these databases script or word are written by different-different writers. In image acquisition a handwritten script image is scanned by scanner and skeleton of document image is detected. Then isolate the script image into words/sub-words and extract feature of individual characters of script. Train the network to recognize handwriting. Perform post-processing to correct the wrongly recognized words.

Handwritten script can be taken either by scanning or by using electronic devices like tablet and digitizer. Based on handwriting input method we can classify handwriting recognition as online or offline.

- **Online HWR:** Machine recognizes the writing when user writes so it is also known as real time or dynamic handwriting recognition. Online handwriting detection requires a transducer to capture handwriting while writing. Devices used are electronic tablet and digitizer which has 200 points/in resolution rate and 100 points/s sampling rate and pen-down or inking sign. Different devices used different technologies which are electromagnetic or electrostatics, acoustic sensing in air medium and pressure sensing for HWR. Electronic tablet came in the late 1950’s which capture the x-y coordinate’s value of movement of pen-tip. Interest in online handwriting recognition increases because of several reasons which are:

1. Growing of hardware component that combine tablet and flat display bring handwritten input and output at the same time or on the same plane. It use electronic ink and provide immediate feedback. Electronic ink show the immediate trace of the motion of stylus tip.

3. **Pre-Processing:** To recognize text of handwritten script image, several document image analysis steps have been performed such as baseline estimation and slope correction, slant correction, smoothening and thinning. Fig. 6 shows the pre-processed handwritten word and detected skeleton [10].
4. **Segmentation:** Handwritten script after preprocessing gives a clean document means it contain sufficient shape information, low noise, high compression. Next step is divide script into word and sub-word. Segmentation is important step because using this step line, word and sub-word of script can be separated.
5. **Feature Extraction:** To avoid complexity of character and increase recognition accuracy feature extraction has been performed. Feature extraction plays an important role in handwriting recognition because it gives the way of representing (characteristics) of a character. For handwriting recognition a set of feature is extracted for each class which helps in distinguishing it from other class.
6. **Training:** To recognize character of handwritten script a training dataset is created which is trained with extracted feature of characters. For training various classifiers can be used such as support vector machine (SVM), self-organizing feature map (SOM), neural network (NN), hidden markov model (HMM) etc.
7. **Post-Processing:** Way to take context information and to correct minor mistakes of handwriting recognition use dictionary. Spell checker for some languages are used in post-processing. N-gram is used to correct statistical error [11].

The problem of handwriting recognition is still challenging and great interest for researchers. Even if human being is well trained optical handwriting reader, their eyes makes about 4% mistakes while reading in the contexts absence [4].

2. To automate office work, it provides a natural way to enter data into machine.
3. It is user friendly.

Disadvantage of online HWR

1. User requires special equipment for HWR.
2. Online HWR captures dynamic or temporal information of the handwriting [7].

Offline HWR: Offline handwriting recognition has been performed on scanned images of script once the writing is completed. Optical recognition of character is part of offline handwriting recognition [7]. Information captured in handwriting recognizing is static and dynamic. Static information consist shape and size of character. Dynamic information's are as follows:

1. Number of strokes
2. Order of stroke.
3. Writing direction of each stroke.
4. Writing speed within strokes.

Challenges in handwriting recognition:

1. Identical representation of some characters such as 0 and O, I and l [7].
2. Cursive way of writing.
3. **Touched character:** Mostly touching character consist only two characters and more than two characters are uncommon. Some touching character interpreted as single character such as when r and n characters are touching each other, it may be misinterpreted as m. Touching character have larger aspect ratio than single character [8].
4. Broken characters
5. Variance in writing style

A. Skeleton Detection: It is first step in handwriting recognition. Noise, slant and stroke in handwritten script make word mismatch and decrease recognition rate. Fig. 7 show handwritten script image and its skeleton. Skeleton detection includes major steps such as Binarization, baseline estimation, slant correction and thinning.

- **Binarization:** Scanned digital image have 0% to 100% intensity. This is primary step to detect skeleton. If intensity of the pixel value of script image is less than a given threshold value, it is set to 0 (black). If intensity is greater than given threshold value, it is set to 1 (white). Once binarization is completed each pixel of image is shown as either black or white color [16].
- **Baseline Estimation:** Height of character is determined by finding important lines, top and bottom which are running alongside characters. Ascender and descended also identified which are used to measure shape of

character. Baseline estimated by drawing horizontal and vertical density histogram as fig. 8 [17].

- **Slant Angle Correction:** Slant of word estimate by evaluating average angle of vertical stroke using edges of stroke. Ending and starting points of line give the slant angle. Edges of stroke determine either by edge detection filter or by finding contour of threshold image. It gives the chain of connected pixel and slant which measure by vertical edges [17]. Partha Pratim Roy et al. have been used Wigner-Ville distribution and vertical projection histogram for slant angle determination and correction. Shear transform is performed then using vertical projection analysis height of the peaks at an angle in interval -45 to +45 degree are estimated [18].
- **Thinning:** Thinning process reduces the amount of information needed to process by removing replica of data points. Most thinning method based on erosion. In erosion over the image, a window is moved and rules are applied. After completing each scan marked points are removed [17]. Soumen Bag proposed a contour based thinning algorithm to find a smooth thinned image. Thinned image of character called skeleton [19].



Figure 7: Selected Image for HWR and its Skeleton [19]

B. Segmentation: After detecting skeleton of script, word segmentation is performed to isolate word and character. Individual character is used for identification. If characters are not correctly segmented, it can't be recognized correctly [18]. It include methods such as junction based segmentation, Zone-wise, region based and component based segmentation.

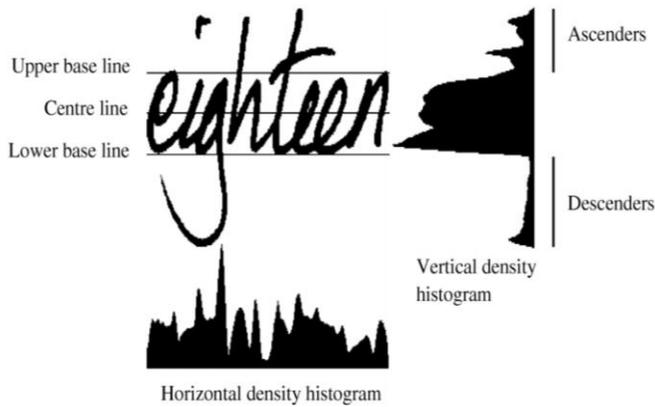


Figure 8: Baseline Estimation [17]

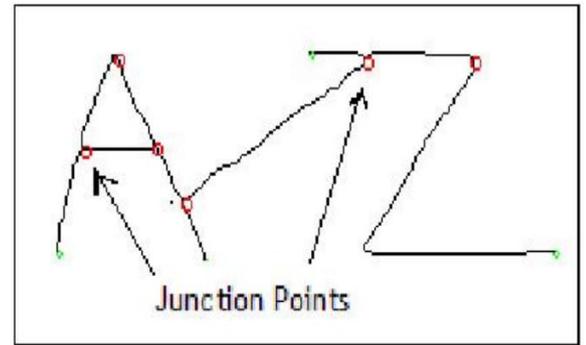


Figure 9: Junction Based Segmentation Method [20]

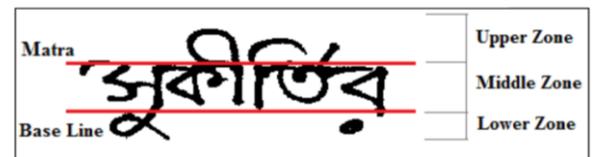


Figure 10: Three Zone of Indic Script (Lower, Middle and Upper Zone) divided by Matra and Baseline [21]

- Junction Based Segmentation:** U.K.S. Jayarathnal and G.E.M.D.C. Bandara presents a new technique for character segmentation on the basis of junction point of joined two digit strings of handwritten script. To isolate the character, handwritten script image divided into individual character. These isolated characters take under the correlation area. Assume that maximum space amid two rows of associated character frame is 50 pixels and space amid two disconnected character frame in a single row is 200 pixels. This isolated area of connected character consists two phase for segmentation, initial and full segmentation phase. To segment connected character skeleton, entered character image identify junction point in the initial segmentation step. Fig. 9 shows the correlation area of text and its junction points. In this step some non-junction points are also found so a rule based segmentation algorithm is applied for complete segmentation [20].
- Zone Wise Segmentation:** A novel method towards Indic handwriting recognition has been proposed by Partha Pratim Roy et al. Indic handwriting detection is tedious because it consist compound character, modifiers, touching and overlapped character etc. At the baseline and upper part, Indic script consist a horizontal line called Matra / Shirorekha as fig. 10. To make a word write 2 or more character side by side. It gives a long line by touching these horizontal lines. Characters hang with the matra. Every Indic writing run from left to right. To display character, several connecting glyphs come out to the left of base character. Horizontally partition the word image into lower, middle and upper zone and detect each zone. Advantage of zone segmentation in Indic script is total number of classes decrease into dissimilar component classes [21].

- Region Based Segmentation:** F. Kurniawan et al. have implemented an intelligent technique of offline cursive words on touching character problem. Fig. 11 shows the region wise segmentation. The script image pixels as feature vector were mapped into coordinate system. Feature vector are partitioned into right, left and middle region of classes. To recognize the character into three regions vertical segmentation has been implemented using self organized feature map (SOM) network which are estimated by finding core zone of word [22].

C. Feature extraction: In feature extraction, characteristics of character are extracted to recognize handwritten script. Each writer has a unique way of writing. Only relevant information is extracted and removes redundant/unnecessary information. Feature extraction can be classified as:

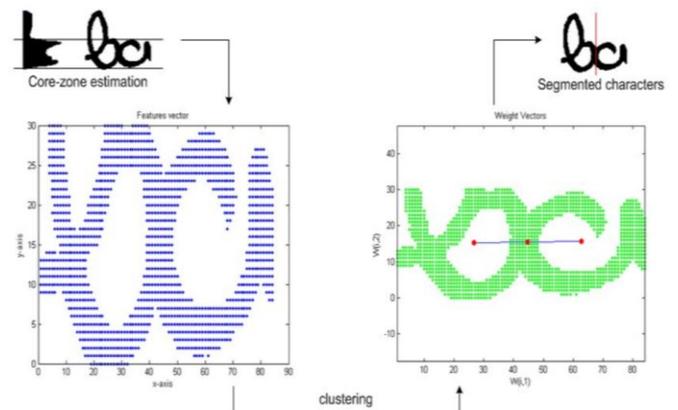


Figure 11: Handwritten Word Recognition using Region Wise Segmentation [22]

- **Structural Feature:** Structural features describe the structural information of script. These are stroke, no. of bifurcation points, contour and no of circles. Structural feature are classify based on geometrical and topological characteristic of character.

1. **Topological features:** These features are used to build a character. Some topological features are cross point, loop, end of line and branch point in the left, right, upper and lower direction, changes occur between two points, hidden dots, curvature between two points, relation and direction of strokes, curve of line, maximum value of point, minima and maxima etc.

2. **Geometrical features:** Many global and local features are considered in geometrical features. These are curvature of word length, fraction of height and width of character bounding box, distance between ending point and y-min vertically, distance between initial and ending point vertically or horizontally, distance covered by two point, words lower and upper mass, stroke width and strokes distance [23]

- **Statistical Feature:** Statistical distribution takes writing style variation. It helps in dimension feature set reduction. It consist density of pixel, mathematical knowledge and moment. Statistical feature based on style of variation at some extent of character. Statistical features are zoning, projection and profile, crossing and distance. [24].

1. **Zoning:** The characters frame is partitioned into numerous zones which are overlapped or non-overlapped. Density and some other feature in these zones are analyzed.

2. **Crossing and Distances:** The frame of character is divided into several regions of specified directions and each zones feature is extracted.

3. **Projections:** Projecting the gray value pixel on the lines of characters in various directions. It forms a 1-D image from 2-D image which is used in character image creation.

- **Pyramid Histogram of Oriented Gradient (PHOG) Feature:** It is a novel sliding window based feature that describes spatial shape features of script by spatial layout and local shape. It consist gradient orientation at each level of pyramid. Each sliding window is separated into cells at each level of pyramid. Pyramid has 4^N different cell at each level of resolution. To get final PHOG, at each level of pyramid concatenate all feature vectors. At level zero L vector shows the histogram L-bins. A cell has $L \cdot 4^N$ structural feature vector so final PHOG descriptor has $L \cdot \sum_{N=0}^K 4^N$ structural feature vector [21].

- D. **Classification:** In this step, script is recognized by using trained network. Artificial neural network (ANN),

support vector machine (SVM) and hidden Markov model (HMM) are used for classification. Combination of these classifiers can be used for classification.

- **Support Vector Machine (SVM):** SVM is a supervised machine learning classification method which is used in pattern recognition application. SVM maps the records into high dimensional input space to construct a hyperplane. Larger margin in hyperplane increase correct classification accuracy. It classifies data into two classes +1 and -1 [25].
- **Artificial Neural Network (ANN):** ANN is a data mining tool that used for pattern recognition and classification. It consist artificial neuron which sends information from one neuron to another neuron. Output of each neurons calculated by sum of inputs. Each neuron associated a weight which is used in learning process. It consist three layers: input, output and hidden layer. Input layer neuron connected to the hidden layer neuron and hidden layer neuron connected to output layer neuron [26].
- **Hidden Markov Model (HMM):** Hidden Markov Model consist many states. It is a finite state model that describes probability distribution over a great number of possible sequences of state. Each state gives out symbols (residues) using probability of symbol-emission and observes sequence of symbols. Probability of state-transition is used to interconnect the states. Continuously states are generated by moving from start/primary state till the end state is reached. HMM contain a large number of continuously ordered states called Markov chain. Sequence of state can't be recognized because states are hidden and selection of state depends on the current state. Only final state/symbol are observed [27].

Problems in hidden Markov model:

- **Scoring Problem:** When hidden Markov model generate an observed sequence, the probability is unknown.
- **Alignment Problem:** Can't observe optimal state sequence to generate the observed sequence.
- **Training Problem:** When a large amount of data is taken. Best structure and parameter for hidden Markov model cannot be calculated.

Contribution of hidden Markov model:

- Hidden Markov model can be trained from aligned and unaligned data.
- HMM used a justifiable statistical treatment for insertion and deletion.

E. Post processing

After performing classification, post processing phase is used. It is integration of additional information/dictionary to improve remove misclassification and recognition accuracy.

Instead of single solution, classification sometimes gives a set of solution. At this time post processing used to select unique solution using high level information. Checking and correction of spelling are the most common post processing operation. Dictionary based error correcting/post processing takes more time and increase with the size of dictionary.

F. Classification Accuracy

Offline handwriting recognition [28] based on artificial neural network done on 2600 samples, written by 100 writers achieve 86.74% recognition rate with minimum time. The hidden Markov model (HMM) and veterbi decoding algorithm is used to decode series of state of HMM [21]. Handwritten English character recognition accuracy is 98.26% accuracy [29]. Arabic handwritten script is recognized using embedded training based on hidden Markov model. It recognized handwritten text with 87.93% accuracy [30]. Support vector machine for recognition of handwritten script on form document using loss function and regularization function. It obtain an accuracy rate of 92.33% for uppercase characters, 98.30% for numeral characters, 83.54% for lowercase characters, and 88.32% for the merger of numeral and uppercase characters [31].

OPTIMAL SYSTEM

As discussed in previous section through various references the segmentation of cursive and running handwritten text is most challenging and burning problem. The methods available are not sufficient enough to segment and recognize the text. We observed that the single length window cannot segment all the characters, so a dynamic length window or combinational window should be used to cover the nearby segmented characters. This should increase the accuracy. Also the single category of features is not enough to recognize the large number of variations among the subjects' handwriting. The handwritings itself has number of variations. So it is better to ask a larger sample from the subject written in the similar conditions for which the system is targeting and having majority of combinations of characters that are used in the language and chance of overlap is there. This also requires a linguistic knowledge to design such training text.

APPLICATIONS OF HANDWRITING RECOGNITION

A. Education System

Education system is a way of learning. It can increase our knowledge and skills. E-learning can save time and printing cost of student [32]. Todd Maurice Uthman et al. proposed a system to evaluate student worksheets. This method takes a worksheet image. Worksheet includes at least one problem for that student gives an answer. Locate answer region and extract marks of student using character recognition techniques. Then store marks as student answer data and assign grade to worksheet [33]. By providing information feedback enables the learning which can be used to increase

future performance. Jorge Villalon et al. proposed a glosser system which enhances feedback for student writing task. Glosser system provides support for teaching and learning of academic writing in English [34].

B. Handwritten Address Interpretation

Handwritten address interpretation is one of the challenging tasks in detecting postal address in mailing services. Detection of correct mailing address is responsible for physically delivering of information or some other things. An address consist characters and digits both which are personal name, street, location, city, state and pin code etc. Handwritten address interpretation involves recognizing of character and digits [35].

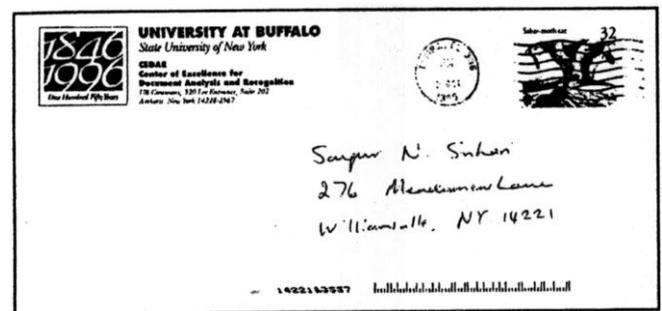


Figure 12: Handwriting Recognition using Remote Computer Reader System [36]

Fig. 12 is the recognition of scanned image of handwritten address on post card. Handwritten address interpretation technique incorporate the letter processing which is mailed by United State Postal Service. Remote computer reader integrated with handwritten address to recognize it [36]. Interpretation result represented as bar code and put at the envelope's bottom then sorting made by the bar code reader [37].

C. Bank Check Recognition

Bank check recognition is a challenging task because bank check has complex and colored background. Bank check contain name of the bank and its identification number, date, amount in words, amount in numbers, symbol of currency and signature. Recognition process involves noise, line removal binarization, segmentation and recognition [2].

CONCLUSION

In this paper we have summarized various terminologies, steps and techniques involved in handwritten character recognition. Firstly we classified the type of problems in this field and then we listed out the steps of the handwriting recognition briefly. We explained all the steps in detail with the study of various techniques used by the researchers. As per our insight the problem still requires increase in the recognition accuracy. We have also summarized outlines for tentative requirements and contents for an optimal system. We have suggested future directions towards an optimal

system. The paper should be beneficial for the researcher looking to work in the field of handwritten character recognition area.

REFERENCES

- [1] JANET EMIG, "Writing as a Mode of Learning", JSTOR (1977), pp. 122-128.
- [2] Rejean Plamondon and Sargur N.Srihari, "On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 1, January 2000.
- [3] R. Plamondon and G. Lorette, "Automatic Signature Verification and Writer Identification-The State of the Art", Pattern Recognition, vol.22, no.2, pp. 107-131, 1989.
- [4] J. Mantas, "An Overview of Character Recognition Methodologies", Pattern Recognition, Vol. 19. No. 6. pp. 425-430, 1986.
- [5] V. K. Govindan and A. P. Shivaprasad, "Character Recognition- A Review", Pattern Recognition, Vol. 23, No. 7, pp. 671-683, 1990.
- [6] Raashid Hussain, AhsenRaza, Imran Siddiqi, Khurram Khurshid and Chawki Djeddi, "A comprehensive survey of handwritten document benchmarks: structure, usage and evaluation", Hussain et al. EURASIP Journal on Image and Video Processing, 2015. DOI 10.1186/s13640-015-0102-5
- [7] Charles C. Tappert, Ching Y. Suen and Toru Wakahara, "The State of the Art in On-Line Handwriting Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence Vol. 12, No. 8, August 1990.
- [8] Tanzila Saba, Ghazali Sulong and Amjad Rehman, "A Survey on Methods and Strategies on Touched Characters Segmentation", International Journal of Research and Reviews in Computer Science (IJRRCS) Vol. 1, No. 2, June 2010.
- [9] M.E. Stevens, "Introduction to the special issue on optical character recognition (OCR)", Pattern Recognition 2, 147-150 (1970).
- [10] Andrew W. Senior and Anthony J. Robinson, "An Off-Line Cursive Handwriting Recognition System", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 20, NO. 3, MARCH 1998.
- [11] Nafiz Arica and Fatos T. Yarman-Vural, "An Overview of Character Recognition Focused on Off-Line Handwriting", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, VOL. 31, NO. 2, MAY 2001.
- [12] U.-V. Marti and H. Bunke, "The IAM-database: an English sentence database for offline handwriting recognition", IJDAR (2002), pp: 39–46.
- [13] Florian Kleber, Stefan Fiel, Markus Diem and Robert Sablatnig, "CVL-Database: An Off-line Database for Writer Retrieval, Writer Identification and Word Spotting".
- [14] Mario Pechwitz, Samia Snoussi Maddouri, Volker Märgner, Noureddine Ellouzeand Hamid Amiri, "IFN/ENIT - Database of Handwritten Arabic Words", CIFED.
- [15] Emmanuele Grosicki and Haikal El Abed, "ICDAR 2009 Handwriting Recognition Competition", 10th International Conference on Document Analysis and Recognition, 2009.
- [16] Amit Choudharya, Rahul Rishib and Savita Ahlawat, "Off-Line Handwritten Character Recognition using Features Extracted from Binarization Technique", AASRI Conference on Intelligent Systems and Control, 2013, pp. 306-312.
- [17] Andrew W. Senior and Anthony J. Robinson, "An Off-Line Cursive Handwriting Recognition System", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 3, March 1998.
- [18] YI LU, "Machine Printed Character Segmentation--An Overview", Pattern Recognition. Vol. 28, No. 1, 1995, pp. 67-80
- [19] Soumen Bag and GauravHarit "An improved contour-based thinning method for character images", Pattern Recognition Letters 32 (2011) 1836–1842.
- [20] U.K.S. Jayarathnal and G.E.M.D.C. Bandara, "New Segmentation Algorithm for Offline Handwritten Connected Character Segmentation", First International Conference on Industrial and Information Systems, ICIIS 2006, 8 - 11 August 2006, Sri Lanka.
- [21] ParthaPratim Roy, Ayan Kumar Bhunia, Ayan Das, PrasenjitDey and Umapada Pal, HMM-based Indic Handwritten Word Recognition using Zone Segmentation, Pattern Recognition, <http://dx.doi.org/10.1016/j.patcog.2016.04.012>
- [22] F. Kurniawan, M. Shafry, D. Daman, A. Rehman, D. Mohamad and S. M. Shamsudin, "Region-Based Touched Character Segmentation in Handwritten Words", International Journal of Innovative Computing, Information and Control Volume 7, Number 6, June 2011.
- [23] Muhammad Arif Mohamad, Dewi Nasien, Haswadi Hassan and Habibollah Haron "A Review on Feature Extraction and Feature Selection for Handwritten Character Recognition", International Journal of Advanced Computer Science and Applications, Vol. 6, No. 2, 2015.

- [24] Muhammad Naeem Ayyaz, Imran Javed and Waqar Mahmood, "Handwritten Character Recognition Using Multiclass SVM Classification with Hybrid Feature Extraction", *Pak. J. Engg. & Appl. Sci.* Vol. 10, Jan., 2012, pp. 57-67.
- [25] J.A.K. Suykens and J. Vandewalle, "Least Squares Support Vector Machine Classifiers", *Neural Processing Letters*, 1999, pp. 293–300.
- [26] David Kriesel, "A Brief Introduction to Neural Networks", 2005.
- [27] Sean R Eddy, "Hidden Markov Models", *Current Opinion in Structural Biology* 1996, pp. 361-365.
- [28] Suman Avdhesh Yadav, Smita Sharma and Shipra Ravi Kumar, "A Robust Approach for Offline English Character Recognition", 1st International conference on futuristic trend in computational analysis and knowledge management (ABLAZE), 2015.
- [29] Rajib Lochan Das, Binod Kumar Prasad and Goutam Sanyal, "HMM based Offline Handwritten Writer Independent English Character Recognition using Global and Local Feature Extraction", *International Journal of Computer Applications (0975 – 8887)* Volume 46– No.10, May 2012.
- [30] Mouhcine Rabi, Mustapha Amrouch and ZouhirMahani, "Recognition of Cursive Arabic Handwritten Text Using Embedded Training Based on Hidden Markov Models", *International Journal of Pattern Recognition and Artificial Intelligence* Vol. 32, No. 1 (2018). DOI: 10.1142/S0218001418600078
- [31] Darmatasia and Mohamad Ivan Fanany, "Handwriting Recognition on Form Document Using Convolutional Neural Network and Support Vector Machines (CNN-SVM)", *Fifth International Conference on Information and Communication Technology (ICoICT)*, 2017.
- [32] Stuart Hepplestone*, Graham Holden, Brian Irwin, Helen J. Parkin and Louise Thorpe, "Using technology to encourage student engagement with feedback: a literature review", *Research in Learning Technology Aquatic Insects* Vol. 19, No. 2, July 2011, pp. 117–127.
- [33] Uthman, Todd Maurice, Frank Porter, and Adam D. Ledgerwood. "System and method for using optical character recognition to evaluate student worksheets." U.S. Patent Application No. 13/012,359.
- [34] Jorge Villalón, Paul Kearney, Rafael A. Calvo and Peter Reimann, "Glosser: Enhanced Feedback for Student Writing Tasks", *de Proceedings of ICALT*, Santander, Spain, 2008.
- [35] Uma Mahadevan and Sargur N. Srihari, "Parsing and Recognition of City, State, and ZIPCodes in Handwritten Addresses" *IEEE*, 1999
- [36] Sargur. N. Srihari and Edward. J. Kuebert, "Integration of Hand-Written Address Interpretation Technology into the United States Postal Service Remote Computer Reader System", *IEEE* 1997.
- [37] Sargur N. Srihari, Venu Govindaraju and Ajay Shekhawat, "Interpretation of Handwritten Address in US Mailstream", *Proceedings Second International Conference on Document Analysis and Recognition*, Tsukuba, Japan, 1993, *IEEE Computer Society Press*, PP. 291-294.