Offline Handwriting Recognition with Emphasis on Character Recognition: A Comprehensive Survey

Aarthna Maheshwari
Information Technology
Mahakal Institute of Technology
Ujjain, India

Khushboo Shah
Information Technology
Mahakal Institute of Technology
Ujjain, India

Abstract—Handwriting has continued to be a means of communication as well as a talisman of an individual. It has also been tool to record information. Machine recognition of handwriting has found its presence in PDA, in portal addresses on envelopes, in amounts in bank checks, in handwritten notes and fields. Character recognition is a process by which computer recognizes letters, numbers or symbols and turn them into digital form. It has gained a lot of use in pattern recognition. It is one of the well liked and challenging area of research

Keywords—Character, Character Recognition, Preproecessing, Segmentation and Classification.

I. INTRODUCTION

A. Nature of Handwriting

An individual's style of writing with hand with the help of writing instruments (Pen or pencil) is known as Handwriting [1]. Handwriting was developed to serve two purposes: a) To enlarge human memory to serve the need to store information permanently and to aid communication since knowledge was transferred from one generation to next verbally, before the advent of handwriting. [1] The communication was facilitated with the help of symbols and rules (to combine) assigned to the language. These symbols are now known as characters which are combined using linguistic rules to form words henceforth

B. History of Handwriting

Before the onset of handwriting, the culture, stories, norms, rituals etc were passed verbally from one generation to the next. With the evolution of cultures around, humans felt the need to standardize communication. The standardization was achieved by using pictographs which was an evolved form of simple drawings, beginning the use of handwriting. [2]

The first systematic handwriting system was "Sumerian" pictographic system, which used clay tablets. This system was then modified into an evolved form, called "Cuneiform" in 3200 B.C. The earliest alphabetical system was developed by "Phoenicians" in eleventh century. This system lacked vowels and consisted on 22 alphabets [3]. Hebrew and Aramaic scripts are heavily influenced by this system.

The survival of handwriting is based on the use of copybooks and writing methods such as Palmer method used to teach handwriting. The reason that handwriting still has an upper hand on digital devices is the convenience and the ease of use as compared to keyboards. Handwriting serves as a talisman and the standard of conformity of an individual [4] and hence can be used for Handwriting Verification.

C. Death of Handwriting

Handwriting facilitated communication using messages as tools formed using linguistic rules. Handwriting manifests credibility since the classification of anything into pre-history and history is based on the presence of handwritten records. However, in recent times, while computer ownership is on the rise, while only 8% of American household owned a computer in 1984, in 2011, 75% households have some form of computing device which certainly shows increased use of typing. [5]

The Washington Post says that cursive handwriting, which was a mandatory part of elementary education, has been disappearing. According to the Common Core Standards in 2011, abolished obligatory teaching of cursive handwriting after the 1st grade.

In August 2013, Netherland witnessed commencement of iPad schools which relied heavily on digital education. It was not mandatory for students to be present in class. Dutch government also gave green signal to the idea. The creator of iPad schools, Dehond, stated that only 4% of the entire coursework relied on handwritten material [7]

D. Handwriting Analysis

The analysis of handwriting (graphoanalysis) analysis physical characteristics and patterns of handwriting with the purpose of identifying the writer and/or to indicate psychological state at the time of writing and involving physical characters of the writer. [8] Graphology is considered a pseudoscience [9] [10] [11] [12] [13]. The four forms of analysis on handwriting are-Handwriting recognition is the task of transforming graphical marks into symbolic representations as understood by the language. For English Orthography, this symbolic representation is the 8 bit ASCII representation of characters. The characters of most written languages of the world can be represented in the form of 16-bit Unicode [14] format. Handwriting Identification is the mechanism of connoting the handwritten text and Handwriting Identification implies determining the author of the handwritten text from a pool of writers, considering each writer's handwriting is unique. Signature verification involves determining whether the signature belongs to a given person or not. Identification and verification are heavily used in forensic science[15]. To determine the nature, character of a specific writer[16] Recognition and interpretation may be used in daily life, example being a pharmacist decodes the medicine written on a prescription. The techniques are primarily used to eradicate variations or possibilities. For applying these techniques the knowledge of subject domain is mandatory

E. Handwriting Input

There are two approaches of providing handwritten input: The first being "Off-line" involves scanning a handwritten text or printed information and converting into digital form. The second called "On-Line" involves writing with a pen shaped device called stylus on an electronic surface called digitizer or a Personal digital assistant having LCD. The strokes (used to form handwritten text) is analysed by a software considering it as electronic ink.

The online handwritten input, the information (two dimensional coordinates of the successive points) is stored as a function of time. In the case of offline input, a static image is used to extract the data (the luminance of the points) .

In terms of storage requirement of raw data, in online system the space requirement is less. The data required for a cursively written in online case is few hundred bytes and in the offline scenario few hundred kilobytes. In the offline case, if a document 8.27x11.69 inch page is scanned at a resolution of 12.8 M (4128 x 3096) results in the scanned image of 1.6 MB. The resolution is the smallest font size that needs reliable recognition, as well as bandwidth required for transmission and storage.

The recognition rate of online recognition system is much higher than offline. For the on-line, unconstrained, handwritten word recognition problem, recognition rates of 95 percent, 85 percent and 78 percent have been recorded for top choice lexicon sizes of 10, 100 and 1000 respectively.[17] In the case of off-line, top choice recognition rate of 80 percent is recorded with a pure cursive words and a 21,000 word lexicon[18].

II. HANDWRITING GENERATION AND PERCEPTION

Handwriting is a learned and practiced skill that involves coordination of various sub-systems of our central-nervous system called motor [19]. The first step in the production of handwriting is at semantic level where the writer intends to write a message. At the lexical and syntactic level, the intended message is transformed into words formed using the lexicons(linguistic symbols) and combined using correct syntax (linguistic rules) . When the individual graphemes are known the writer selects specific allographs. Below the level, the allographs are transformed into movement patterns of our hand.

There are two models of handwriting modeling: Bottom-up and Top-down[20]. Bottom-up features focus on features of human hand wring such as slant, pressure, velocity and tries to reproduce it with observation. The top-down approach focus on the psychological aspect of handwriting such as motor learning, motor movement, planning etc.

III. OFFLINE HANDWRITING RECOGNITION

In offline handwriting recognition, the data to be recognized have been scanned and stored as an image. The system relies on prior knowledge of the domain, where task specific constraints are available. To solve the limitations of offline handwriting recognition, several system have been proposed. A complete overview of the early work has been referenced from [Bunke (2003); Plamondon and Srihari (2000); Steinherz et al. (1999); Suen et al. (2000); Vinciarelli (2002)].[23]

[K.Sirlantzis and Hoque (2001)] proposed a multiple classifier system trained on Freeman chain-code representation of the character contours combined with sn tuple classifiers which increased the recognition rate. Gunter and Bunke (2004b performed an in depth investigation on the variability of quantum of training set on the classification of information using multiple classifier system. All these classifier had some constarints for classifiers, to overcome this [Bertolami & Bunke (2005)] proposed a model using multiple classifiers for an unconstrained handwritten line recognition. Algorithms were then developed for each stage of recognition to achieve precision. For preprocessing to imporve the image, an algorithm to normalize the intensity of background light using adaptive linear function was proposed. This algorithm was based on approximation and was primarily used for recognition of historic data. [59]

Other stages such as normalization witnessed methods to correct structural properties of handwritten text using gradient orientation of the digitized image. Slant and skew can be corrected using this method. [60]

A. Comparability of Recognition Rates

A number of studies have been published whose recognition rates range from 50 to even a perfect 100% (in certain conditions). However a perfect system does not exist. For example, a postal address pertaining to a country can be recognized using the knowledge of the pincodes and improve the accuracy. But for that a certain level of pre-processing is required in the image available.

Three factors which govern the recognition and affect the comparability of the system are: the considered recognition task, which determines the overall complexity of the problem, the data set which may differ in terms of quality and quantum. And the quantity of data used to train and test the system.

The performance of a system crucially depends on the considered task. For example, in the case of isolated digit recognition the performance is usually higher than in unconstrained handwritten text line recognition. When comparing recognition results, one must be aware of the following aspects:

- *Number of classes*: With an increasing number of classes the task becomes more difficult. In character recognition, for example, it is assumed that recognizing ten different digits produces higher classification results than recognizing all the 25 × 2 letters of the Latin alphabet.
- *Input (isolated vs. sequence):* The task becomes easier if the boundaries of the characters are known, in which case a recognition system for isolated characters can be applied instead of a word recognizer.
- *Vocabulary size:* As a rule of thumb, a higher performance can be expected for smaller vocabulary sizes under the constraint that the vocabulary covers all data in the test set.
- *Number of writers*: The most difficult task is writer-independent recognition, i.e., when there is no training data available of the writers represented in the test set. For writer-dependent recognition it must be considered whether there is one recognition system for only one writer or for a number of writers.
- *Language model:* Recognition systems can gain additional information from statistical language models. One can expect higher recognition rates from systems utilizing a language model.

The other important aspect is data set. A publically available data set is generally used to train and test the system because if not used, it is impossible to make a direct comparison of the results. The UNIPEN database [Guyon et al. (1994)] is a large online handwriting database. It contains mostly isolated characters, single words, and a few sentences on several topics..

B. Pre-processing

A data has to be operated to be able to recognize hence several functions such as converting into binary form, noise removal to enhance an image, segmentation of line and words and isolation of individual characters is required [24

C. Thresholding

The task of thresholding is to isolate foreground (ink) and background (paper)[25]. The histogram of the image consists of 2 peaks (for grayscale image): a higher peak for white (paper) and lower peak for black (for ink). Thresholding determines the optimal value in the valley of two peaks [26]

D. Noise Removal

Noise removal is a major concern for both offline and online handwritten character recognition. Digitizing an image may introduce noise in the data. The transmission of information is also a factor in introducing noise. Smoothing operations are used to introduce errors during capturing of image.

Thinning algorithms are used to convert offline-data into on-line data. Unfortunately, thinning introduce errors and cannot be relied on.

E. Line Segmentation

There are various approaches to convert handwritten text into lines, words and characters. It can be accomplished by observing the horizontal histogram profile at certain ske angles [34]. A method uses imaginary line between two consecutive lines of handwritten text on which people write which is calculated using approximation of local maxima and minima from each component (stroke) along with clustering to club maximas and minimas. [35]

F. Word and character segmentation

After line separation, next is word separation. He methods for word separation look for physical gaps between the strokes [36[37]. The method relies on the concept of gap comparison as gap between letters in less than gap between words. Other method [38] uses variation in gaps between the words.

Bifurcation of words and characters is accomplished by ligatures and concavities [39] features. Various features of handwriting like height of character, space between characters is used for e purpose.

IV. CHARACTER RECOGNITION

Character recognition deals with classifying the characters of digitized image into symbolic class. For English Orthography, there are 4 classes: Upper case, Lower case, Digits and special symbols. Te character is determined by extracting the shape of he character. The fastest method is ANN and most accurate method is nearest neighbor.

Recognition of handwritten characters is difficult than single, machine printed characters. A survey of character segmentation can be found in [45]. In some cases models are required to constrain the choices.

- *Matrix Matching:* Matrix Matching converts each character into a pattern with a matrix, and then compares the pattern within a matrix, and then compares the pattern with an index of known characters. Its recognition is strongest on monotype and uniform single column pages.
- *Fuzzy Logic:* Fuzzy logic is a multiple valued logic that allows intermediate values to be defined between conventional evaluations like yes/no, true/false, 0/1 etc. It is used whn uncertainty is involved.
- *Feature Extraction:* The method defines each each character by the presence or absence of certain features, including height, width, density, loops, lines etc It is perfect for OCR of magazines, laser print and high quality of images.
- *Structural Analysis:* It examines and identifies characters by examining their sub feature shape of their image, sub-vertical and horizontal histograms. Its character repair capability is great for low quality text and newsprints.
- *Neural Networks:* This strategy simulate s ten way human neural network system works. It samples the pixels in each image and matches them to a known index of pixel pattern. The ability to recognize characters through abstraction is great for faxe

ISSN No:-2456 –2165

document and damaged text. It is ideal for processing stock market data or finding trends in graphical patterns.

The main approaches of offline handwritten word recognition can be divided into two classes: holistic and work Extensive survey on isolated handwritten character recognition can be found in [46], [47], [48],[49].

V. AAPLICATIONS OF OFFLINE HANDWRITING RECOGNITION

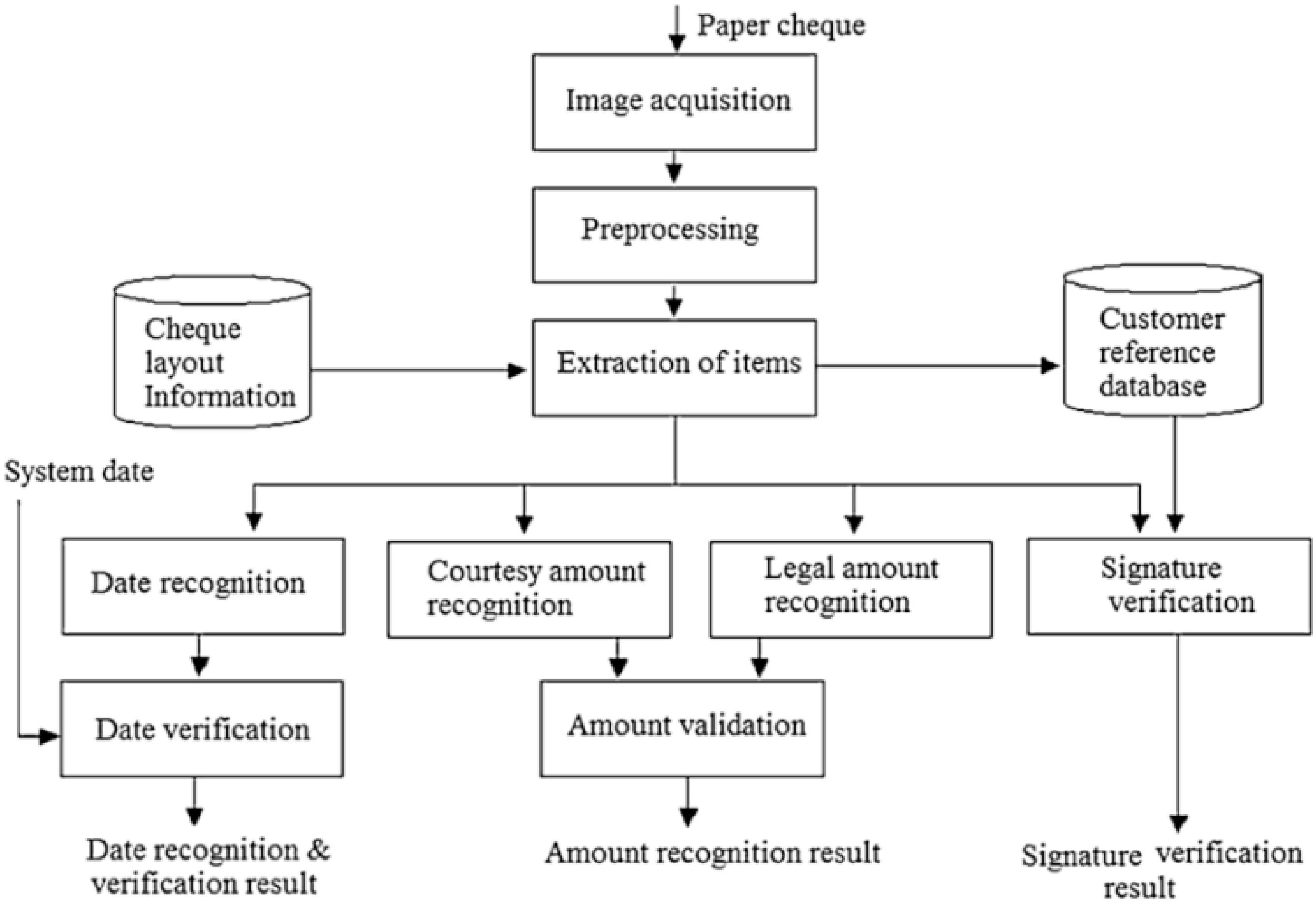
The most prominent applications of offline handwriting recognition are in reading postal address interpretation, bank address and forms.

A. Handwritten Address Interpretation

Handwritten address interpretation is used to classify a letter based on the location. The address consists of Name, country, state, city, pincode, street address and Phone number. [57], [58]. The handwritten address interpretation uses database as a source of information i.e. collection of all the state, city, pincode information. A physical implementation of the system has been installed in United States Postal Service.

B. Bank Cheque Recognition

Bank cheque recognition system consists of various operations such as machine printed numeral recognition, signature verification, courtesy amount recognition, legal amount recognition; The first step is image acquisition. The image is acquired by scanner. Next step in Machine printed numeral recognition involving bank identification code, bank agency identification code, check number and customer account number. The courtesy amount field, legal amount field and signature field is used to remove background and remove guidelines. The courtesy amount recognition module recognizes numerals by hypothesis and verification. The legal amount recognition module includes 3 recognizers. Amount validation module accepts amount, if it is greater than a threshold. Signature verification uses signature image and uses the information from the database of users.



VI. CONCLUSION

This paper addresses the handwriting, its history, current state of art, various uses of it in depth. This paper also highlights generation of handwriting in depth. This paper surveys character recognition, the methods developed till date in brief and practical utility and recognition of those methods. This paper also brings into limelight the practical implementation of offline handwritten character recognition implemented in USA, and its limitations with a scope to improve the recognition rates in the field.

REFERENCES

[1]. https://en.wikipedia.org/wiki/Penmanship

[2]. https://www.vletter.com/help/font-faq/history-of-handwriting.html

[3]. Robinson, Andrew (2007). The Story of Writing. London: Thames & Hudson Ltd. p. 165.

[4]. T.P. Thornton, Handwriting in America: A cultural Hsotory. Yale Univ, 1996

[5]. Linsey Hopman, Cursive over typeface: the importance of teaching handwriting instead of typing, The University of Leiden 2014

[6]. http://www.literacytrust.org.uk/media/5371_children_s_onscreen_reading_overtakes_reading_in_print

[7]. Handwriting is beneficial to Children's Cognitive Development', FYI Living, 22 February 2011 < http://fyiliving.com/research/handwriting-is-beneficial-to-childrenscognitive-development/> (21 May 2014).

[8]. Longman Dictionary of Psychology and Psychiatry, Longman Group United Kingdom, 1983.

[9]. Nevo, B Scientific Aspects Of Graphology: A Handbook Springfield, IL: Thomas: 1986

[10]. ^ Jump up to:[a] [b] "Barry Beyerstein Q&A". Ask the Scientists. Scientific American Frontiers. Retrieved 2008-02-22. "they simply interpret the way we form these various features on the page in much the same way ancient oracles interpreted the entrails of oxen or smoke in the air. i.e., it's a kind of magical divination or fortune telling where 'like begets like'".

[11]. ^ Jump up to:[a] [b] BEYOND SCIENCE? Paper Personality, on season 8 , episode 2 of Scientific American Frontiers.

[12]. ^ Jump up to:[a] [b] James, Barry (3 August 1993). "Graphology Is Serious Business in France : You Are What You Write?". New York Times. Retrieved 18 September 2010.

[13]. ^ Jump up to:[a] [b] [c] [d] Dunning, Brian. "Skeptoid #363: All About Graphology". Skeptoid. Retrieved2 September 2016.

[14]. “The Unicode consortium”, The Unicode Standard, Worldwide

[15]. R. Plamondon, D. Lprosti, I,R,B.Schomaker and R. Srihari, “On-line Handwriting Recognition-The State of the art”,Pattern Recignoiton, Col. 22, no. 2, pp.107-131, 1989

[16]. V. Bouletreau N. Vincent R. Sabourin H. Emptoz "Handwriting and Signature: One or Two Personality Identifiers?"Proc. 14th Int'l Conf. Pattern Recognition, pp. 1 758-1 760 1998-Aug.

[17]. S.N. Srihari "High Performance Reading Machines" Proc. IEEE vol. 80 no. 7 pp. 1 120-1 132 1992.

[18]. G. Seni R.K. Srihari N. Nasrabadi "Large Vocabulary Recognition of On-Line Handwritten Cursive Words" IEEE Trans. Pattern Analysis and Machine Intelligence vol. 18 no. 7 pp. 757-762 July 1996.

[19]. Garipelli Gangadhar, Denny Joseph, V. Srinivasa Chakravarthy, An oscillatory neuromotor model of handwriting generation, International Journal of Document Analysis and Recognition (IJDAR), 2007

[20]. R. Plamondon F.J. Maarse "An Evaluation of Motor Models of Handwriting" IEEE Trans. Systems Man and Cybernetics vol. 19 no. 5 pp. 1 060-1 072 1989.

[21]. Hollerbach, J.M. Biol. Cybern. (1981) 39: 139. https://doi.org/10.1007/BF00336740

[22]. Jeffery J Summers, Approaches to study of motor control and learning, University of Melbourne

[23]. Bunke Horst, Liwicki Marcus, Recognition Of Whiteboard Notes: Online, Offline And Combination, Series in Machine Perception and Artificial Intelligence: Volume 71

[24]. M. Cheriet C.Y. Suen "Extraction of Key Letters Script Recognition" Pattern Recognition Letters vol. 14 pp. 1 009-1 017 1993.

[25]. P.K. Sahoo S. Soltani A.K.C. Wong Y.C. Chen "A Survey of Thresholding Techniques" Computer Vision Graphics and Image Processingvol. 41 pp. 233-260 1988.

[26]. P.W. Palumbo P. Swaminathan S.N. Srihari "Document Image Binarization: Evaluation of Algorithms" Proc. SPIE Applications of Digital Image Processing IX vol. 697 pp. 278-285 1986-Aug.

[27]. Y. Liu S.N. Srihari "Document Image Binarization Based on Texture Features" IEEE Trans. Pattern Analysis and Machine Intelligence vol. 19 no. 5 pp. 1-5 May 1997.

[28]. N. Otsu "A Threshold Selection Method from Gray-Scale Histogram" IEEE Trans. Systems Man and Cybernetics vol. 8 pp. 62-66 1978.

[29]. Z. Shi V. Govindaraju "Character Image Enhancement by Selective Region Growing" Pattern Recognition Letters vol. 17 pp. 523-527 1996.

[30]. Y. Nakajima S. Mori S. Takegami S. Sato "Global Methods for Stroke Segmentation" Int'l J. Document Analysis and Recognition vol. 2 pp. 19-23 1999.

[31]. S.N. Srihari V. Govindaraju "Separating Handwritten Text from Overlapping Contextual Contours" Proc. Second Int'l Workshop on Handwriting Recognition (IWFHRII) pp. 111-119 1991-Sept.

[32]. L. Lam S.W. Lee C.Y. Suen "Thinning Methodologies: A Comprehensive Survey" IEEE Trans. Pattern Analysis and Machine Intelligence vol. 14 pp. 869-885 1992.

[33]. R. Plamondon C.Y. Suen M. Bourdeau C. Barrire "Methodologies for Evaluating Thinning Algorithms for Character Recognition" Int'l J. Pattern Recognition and Artificial Intelligence vol. 7 no. 5 pp. 1 247-1 270 1993.

[34]. S.N. Srihari V. Govindaraju "Analysis of Textual Images Using the Hough Transform" Machine Vision and Applications vol. 2 pp. 141-153 1989.

[35]. G. Kim V. Govindaraju S.N. Srihari "An Architecture for Handwritten Text Recognition Systems" Int'l J. Document Analysis and Recognition vol. 2 pp. 37-44 1999.

[36]. U. Mahadevan R.C. Nagabhushanam "Gap Metrics for Word Separation in Handwritten Lines" Proc. Third Int'l Conf. Document Analysis and Recognition pp. 124-127 1995-Aug.

[37]. G. Seni E. Cohen "External Word Segmentation of Off-Line Handwritten Text Lines" Pattern Recognition vol. 27 no. 1 pp. 41-52 1994.

[38]. G. Kim V. Govindaraju "Handwritten Phrase Recognition as Applied to Street Name Images" Pattern Recognition vol. 31 no. 1 pp. 41-51 1998.

[39]. G. Kim V. Govindaraju "A Lexicon Driven Approach to Handwritten Word Recognition for Real Time Applications" IEEE Trans. Pattern Analysis and Machine Intelligence vol. 19 no. 4 pp. 366-379 Apr. 1997.

[40]. J. Franke L. Lam R. Legault C. Nadal C.Y. Suen "Experiments with CENPARMI Data Base Combining Different Classification Approaches" Proc. Third Int'l Workshop Frontiers of Handwriting Recognition (IWFHRIII) pp. 305-311 1995-May.

[41]. T.K. Ho J.J. Hull S.N. Srihari "Decision Combination in Multiple Classifiers Systems" IEEE Trans. Pattern Analysis and Machine Intelligence vol. 16 no. 1 pp. 66-75 Jan. 1994.

[42]. Y.S. Huang C.Y. Suen "The Behavior-Knowledge Space Method for the Combination of Multiple Classifiers" Proc. IEEE Conf. Computer Vision Pattern Recognition pp. 347-352 1993.

[43]. J. Kittler M. Hatef R.P.W. Duin "Combining Classifiers" Proc. 13th Int'l Conf. Pattern Recognition pp. 897-901 1996-Aug.

[44]. R.K. Powalka N. Sherkat R.J. Whitrow "Multiple Recognizer Combination Topologies" Handwriting and Drawing Research: Basic and Applied Issues. pp. 329-342 1996.

[45]. R. Casey E. Lecolinet "A Survey of Methods in Strategies in Character Segmentation" IEEE Trans. Pattern Analysis and Machine Intelligence vol. 18 pp. 690-706 1996.

[46]. A. Amin "Off-Line Character Recognition: A Survey" Proc. Fourth Int'l Conf. Documents Analysis and Recognition (ICDAR '97) pp. 596-599 1997-Aug.

[47]. V.K. Govindan A.P. Shivaprasad "Character Recognition: A Review" Pattern Recognition, vol. 23 no. 7 pp. 671-683 1990.

[48]. S. Mori K. Yamamoto M. Yasuda "Research on Machine Recognition of Handprinted Characters", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 6 no. 4 pp. 386-405 Apr. 1984.

[49]. C.Y. Suen M. Berthod S. Mori "Automatic Recognition of Handprinted Characters: The State of the Art" Proc. IEEE, vol. 68 pp. 469-487 1980.

[50]. J.C. Simon "Off-Line Cursive Word Recognition" Proc. IEEE, vol. 80 no. 7 pp. 1 150-1 160 1992.

[51]. Y.S. Huang C.Y. Suen "The Behavior-Knowledge Space Method for the Combination of Multiple Classifiers" Proc. IEEE Conf. Computer Vision Pattern Recognition, pp. 347-352 1993.

[52]. R. Plamondon F.J. Maarse "An Evaluation of Motor Models of Handwriting" IEEE Trans. Systems Man and Cybernetics vol. 19 no. 5 pp. 1 060-1 072 1989.

[53]. M. Gilloux J.M. Bertille M. Leroux "Recognition of Handwritten Words in a Limited Dynamic Vocabulary" Proc. Third Int'l Workshop Frontiers of Handwriting Recognition (IWFHRIII) pp. 417-422 1995-May.

[54]. R. Bozinovic S.N. Srihari "Off-Line Cursive Script Recognition" IEEE Trans. Pattern Analysis and Machine Intelligence vol. 11 no. 1 pp. 68-83 1989.

[55]. M. Eden "Handwriting and Pattern Recognition" IRE Trans. Information Theory, vol. 8 1962.

[56]. P.D. Gader M. Mohammed J.H. Chiang "Handwritten Word Recognition with Character and Inter Character Neural Networks" IEEE Trans. System Man and Cybernetics, vol. 27 no. 1 pp. 158-164 1997.

[57]. E. Cohen J.J. Hull S.N. Srihari "Understanding Handwritten Text in a Structured Environment: Determining ZIP Codes from Addresses" Int'l J. Pattern Recognition and Artificial Intelligence, vol. 5 pp. 221-264 1991.

[58]. A.C. Downton R.W.S. Tregidgo C.G. Leedham Hendrawan "Recognition of Handwritten British Postal Addresses" From Pixels to Features III, pp. 129-144 1992.

[59]. Z. Shi, V. Govindaraju "Historical document image enhancement using background light intensity normalization, Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on",26 Aug. 2004 .

[60]. Changming Sun, Deyi Si, "Skew and slant correction for document images using gradient direction", Document Analysis and Recognition, 1997., Proceedings of the Fourth International Conference on, 8-20 Aug. 1997.