AN OVERVIEW AND APPLICATIONS OF OPTICAL CHARACTER RECOGNITION

Ali Mir Arif Mir Asif [1], Shaikh Abdul Hannan[2], Yusuf Perwej[3], Mane Arjun Vithalrao[4]

[1]Institute of Management Studies & I.T., Aurangabad (M.S.)(India)
[2,3]Department of CS & IT, Albaha University, Albaha, (Saudi Arabia)
[4]MGM's Dr. G. Y. Pathrikars College of C. S. & I.T., Aurangabad (M.S.) (India)

ABSTRACT

Optical character recognition, usually abbreviated to OCR, is the mechanical or electronic conversion of scanned or photographed images of typewritten or printed text into machine-encoded/computer-readable text. It is widely used as a form of data entry from some sort of original paper data source, whether passport documents, invoices, bank statement, receipts, business card, mail, or any number of printed records. It is a common method of digitizing printed texts so that they can be electronically edited, searched, stored more compactly, displayed on-line, and used in machine processes such as machine translation, text-to-speech, key data extraction and text mining. OCR is a field of research in pattern recognition, artificial intelligence and computer vision. Optical Character Recognition or OCR is the electronic translation of handwritten, typewritten or printed text into machine translated images. It is widely used to recognize and search text from electronic documents or to publish the text on a website [1]. A large number of research papers and reports have already been published on this topic. The paper presents introduction, major research work and applications of Optical Character Recognition in various fields. At the first introduction of OCR will be discussed and then some points will be stressed on the major research works that have made a great impact in character recognition. And finally the most important applications of OCR will be covered and then conclusion.

Keywords: Calligraphy, Hand Written Characters, Optical Character Recognition (OCR), Practical Applications, Text.

I INTRODUCTION

The paper presents introduction, major research work and applications of Optical Character Recognition in various fields. At the first introduction of OCR will be discussed and then some points will be stressed on the major research works that have made a great impact in character recognition. And finally the most important applications of OCR will be covered and then conclusion. OCR is generally an "offline" process, which analyzes a static document. Handwriting movement analysis can be used as input to handwriting recognition [2]. Instead

of merely using the shapes of glyphs and words, this technique is able to capture motions, such as the order in which segments are drawn, the direction, and the pattern of putting the pen down and lifting it. This additional information can make the end-to-end process more accurate. This technology is also known as "on-line character recognition", "dynamic character recognition", "real-time character recognition", and "intelligent character recognition". Character recognition techniques associate a symbolic identity with the image of a character. This problem of replication of human functions by machines (computers) involves the recognition of both machine printed and handprinted/cursive-written characters [3]. Highlight in 1950's [4], applied throughout the spectrum of industries resulting into revolutionizing the document management process. Optical Character Recognition or OCR has enabled scanned documents to become more than just image files, turning into fully searchable documents with text content recognized by computers. Optical Character Recognition extracts the relevant information and automatically enters it into electronic database instead of the conventional way of manually retyping the text. Optical Character Recognition is a vast field with a number of varied applications such as invoice imaging, legal industry, banking, health care industry, etc. OCR is also widely used in many other fields like Captcha, Institutional repositories and digital libraries, Optical Music Recognition without any human correction or human effort, Automatic number plate recognition and Handwritten Recognition [5]-[10].

Early optical character recognition could be traced to activity around two issues: expanding telegraphy and creating reading devices for the blind. In 1914, Emanuel Goldberg developed a machine that read characters and converted them into standard telegraph code. Around the same time, Edmund Fournier d'Albe developed the Optophone, a handheld scanner that when moved across a printed page, produced tones that corresponded to specific letters or characters. In the late 1920s and into the 1930s Emanuel Goldberg developed what he called a "Statistical Machine" for searching microfilm archives using an optical code recognition system. In 1931 he was granted USA Patent number 1,838,389 for the invention. The patent was acquired by IBM. In 1974, Ray Kurzweil started the company Kurzweil Computer Products, Inc. and continued development of omni-font OCR, which could recognize text printed in virtually any font (Kurzweil is often credited with inventing omni-font OCR, but it was in use by companies, including CompuScan, in the late 1960s and 1970s [11]. Kurzweil decided that the best application of this technology would be to create a reading machine for the blind, which would allow blind people to have a computer read text to them out loud.

This device required the invention of two enabling technologies the CCD flatbed scanner and the text-to-speech synthesizer. On January 13, 1976, the successful finished product was unveiled during a widely reported news conference headed by Kurzweil and the leaders of the National Federation of the Blind. In 1978, Kurzweil Computer Products began selling a commercial version of the optical character recognition computer program. LexisNexis was one of the first customers, and bought the program to upload paper legal and news documents onto its nascent online databases [12]. Two years later, Kurzweil sold his company to Xerox, which had an interest in further commercializing paper-to-computer text conversion. Xerox eventually spun it off as Scansoft, which merged with Nuance Communications. In the 2000s, OCR has been made available online as a service (WebOCR), in a cloud computing environment, and in mobile applications like real-time translation of foreign-

language signs on a smartphone. Various commercial and open source OCR systems are available for most common writing systems, including Latin, Cyrillic, Arabic, Hebrew, Indic, Bengali (Bangla), Devanagari, Tamil, Chinese, Japanese, and Korean characters. The origins of character recognition [13]-[15] can be found in 1870 when Carey invented the retina scanner, that is an image transmission system using a mosaic of photocells, and later in 1890 when Nipkow invented the sequential scanner which was a major break-through both for modern television and reading machines. However, character recognition first appeared as an aid to the visually handicapped and the first successful attempts were made by the Russian scientist Tyurin in 1900. The next attempts that have been reported are the Fourier d'Albe's Optophone of 1912 and Thomas' tactile "relief" device of 1926.

The modern version of OCR [16] appeared in the middle l940s with the development of the digital computer. For the first time, OCR was realised as a data processing approach, with particular application to the business world. From that perspective, David Shepard, founder of the Intelligent Machine Research Co. can be considered as the pioneer of the development and building of commercial OCR equipment. Character recognition is better known as optical character recognition (OCR) since it deals with recognition of optically processed characters rather than magnetically processed [17] ones. Though the origin of character recognition can be found as early as 1870, it first appeared as an aid to the visually handicapped, and the first successful attempt was made by the Russian scientist Tyurin in 1900 [18]. The modern version of OCR appeared in the middle of the 1940s with the development of the digital computers. Thenceforth it was realized as a data processing approach with application to the business world. The principal motivation for the development of OCR systems is the need to cope with the enormous flood of paper such as bank cheques, commercial forms, government records, credit card imprints and mail sorting generated by the expanding technological society. OCR machines have been commercially available since the middle of the 1950s. Since then extensive research has been carried out and a large number of technical papers and reports have been published by various researchers in the area of character recognition. Several books have been published on optical character recognition [19]-[27]. Also special issues and reports on the topic have repeatedly appeared in the proceedings of the International Joint Conferences on Pattern Recognition and of the International System, Man and Cybernetics Conferences- Research works also appear in various other Conferences such as British Conferences on Pattern Recognition, and The Scandinavian Conferences on Image Analysis. State of the art reports on character recognition (research have been presented by Nagy [28], Harmon [29]), Stallings [30] Suen et al. [31), Mori et al. [32] Mantas [18]), Davis and Yall [33], and Chatterji [34].

II MAJOR RESEARCH WORK IN OCR

A notable early attempt in the area of character recognition research is by Grimsdale et al. [35] in 1958. In their method, the input character pattern obtained by a flying spot scanner is described in terms of length and slope of straight line segments and length and curvature of curved segments. The description is compared with that of the prototype stored in the computer in order to reach the proper decision about the identity of the unknown character. Another important work is the analysis-by-synthesis method suggested by Eden [36], [37] at M.I.T.

He put forward the idea that all Latin script characters can be formed by 18 strokes, which in turn can be generated from a subset of 4 strokes, namely, hump, bar, hook, and loop. Some of the examples of the works in this directions are those by Blesser et al. [38], Cox et al. [39], Shillman et al. [40], Yoshida and Eden [41] and Berthod [42]. Blesser et al. proposed a theoretical approach based on phenomenological attributes. Cox et al. presented two main groups of grammar-like rules to deal with variability in type fonts. Three experimental techniques for studying ambiguous characters and for investigating relationship between physical and functional attributes were suggested by Shillman et al. Yoshida and Eden proposed a Chinese character recognition system which employs a generative process to extract a stroke sequence from the input pattern, and a look up dictionary of strokes to effect recognition. Berthod utilized Eden's primitives for cursive script analysis.

In the sixties, Narasimhan suggested labeling schemata for syntactic description of pictures [43], and a syntax directed interpretation of classes of pictures [44]. In another work [45], he proposed a recognition technique based on description and generation. Using primitives and relations, he described a specification language for handprinted FORTRAN character recognition. Later, Narasimhan and Reddy [46] put forward a syntax-aided recognition scheme, wherein they incorporated in the decision rule some flexibility required for the satisfactory performance of a recognition system. The authors expressed the views that the rule currently in use must be refined, modified, and augmented continuously on the basis of the experience and other relevant knowledge acquired. Pavlidis and Ali [47] and Ali and Pavlidis [49] utilized split-and-merge algorithm [48] for the polygonal approximation of characters for numeral recognition. A feature generation technique for syntactic pattern recognition by approximating character boundary by polygons and then decomposing on the basis of concavity is suggested by Feng and Pavlidis [50] in 1974. Major research activities in character recognition are now centered about the recognition of handprinted Chinese characters, which was once considered to be a very hard problem and regarded as one of the ultimate goals of character recognition research. In 1966, Casey and Nagy [51] at IBM presented one of the first attempts at Chinese character recognition.

In late 1970, Agui and Nagahashiu [52] suggested a description method for handprinted Chinese character recognition. In their technique, a Chinese character is represented by partial patterns using three relations, namely concatenate, cross and near. The relations of relative location among partial patterns are used for categorization of the partial patterns. In 1980, Arakawa [53] suggested an on-line hand- written character recognition system for Japanese characters. Fourier coefficients of pen-point movement loci relating to strokes are utilized as feature vectors. A modified relaxation technique, incorporating the knowledge about the Chinese characters into the training system to reduce computational load is suggested by Leung et al. [54]. Finally Yong [55] suggested recognition via neural networks for achieving fast recognition of handprinted Chinese characters. Not many attempts have been carried out on the recognition of Indian character sets. However, some major works are reported on Devanagari (an Indian script used for writing Sanskrit, Hindi and some other languages [56-60] and Tamil [61-65] character recognition. Some attempts are also reported on Brahmi (a script widely used all over India during third century BC) [62], Telugu [66] and Bengali [67] characters. Some of the important character recognition research of the early eighties are those by Tanaka et al. [68], Sarvarayudu and Sethi [69], Shridhar and Badreldin [70] [71]; Sato et al. [72] and Evangelisti [73]. A brief description of them is given in the following.

A large amount of research work has been carried out in the mid eighties and after. A few of them are reviewed here. They include contextual post processing by Nagy et al. [74] and Sinha [60] word/script recognition by Almuallim and Yamaguchi [75], El-sheikh and Guindi [76], Hull [77], Aoki and Yamaya [78] Wong and Fallside [79], and Shrihari and Bozinovic [80], separation of connected characters by Tampi and Chetlur [81] and Ting and Ward [82], numeral recognition by Lam and Suen [83], and Baptista and Kulkarni [84], multifont learning by Cannat et al. [85] [86], learning by experience by Malyan and Sunthankar [87] Pitman's shorthand recognition by Leedham and Downton [88] 89], pattern description and generation technique by Nagahashi and Nakatsuyama [90], description aided recognition by Harjinder [91] and pre-classification and recognition using Walsh transform by Huang and Lung [92]. Other important works include the work by Wolberg [93] who suggested a syntactic omni-font system that recognizes a wide range of fonts including hand-printed characters; on performance testing of mixed font variable size character recognizers by Lam and Baird [94], about the vectorizer and feature extractor for the document reader suggested by Pavlidis [95], and on the guide lines for designing feature vectors for use with large character sets given by Hagita and Masuda [96]. The scopes of all the above attempts were limited because they use simple features which do not exactly or directly reflect the structural details of the characters. They cannot represent the varying structural complexities of different alphabet sets. Moreover, with such simple features the automatic design problem will be easier to handle. Recently, the authors [97] have suggested an automated approach to the design of recognizers suitable for structurally different character sets. The approach is somewhat similar to that of Kami's [98]. However, a flexible and unified/general feature representation is employed to take care of the controlled incorporation of structural details (to describe various character classes) depending upon the complexity of an alphabet set.

III APPLICATIONS OF OCR

Optical character recognition has been applied to a number of applications. Some of the literatures covering these are languages other than English, namely, Latin, Cyrillic, Arabic, Hebrew, Indic, Bengali (Bangla), Devanagari, Tamil, Chinese, Japanese, Korean, etc. Highlight in 1950’s [99], applied throughout the spectrum of industries resulting into revolutionizing the document management process. Optical Character Recognition or OCR has enabled scanned documents to become more than just image files, turning into fully searchable documents with text content recognized by computers. Optical Character Recognition extracts the relevant information and automatically enters it into electronic database instead of the conventional way of manually retyping the text. Optical Character Recognition is a vast field with a number of varied applications such as Practical applications, Invoice imaging, Legal industry, Banking, Healthcare, etc. OCR is also widely used in many other fields like Captcha, Institutional repositories and digital libraries, Optical Music Recognition without any human correction or human effort, Automatic number plate recognition, Handwritten Recognition and Other Industries. Some of them have been explained below:

3.1 Practical Applications

In recent years, OCR (Optical Character Recognition) technology has been applied throughout the entire spectrum of industries, revolutionizing the document management process. OCR has enabled scanned documents

to become more than just image files, turning into fully searchable documents [100] with text content that is recognized by computers. With the help of OCR, people no longer need to manually retype important documents when entering them into electronic databases. Instead, OCR extracts relevant information and enters it automatically. The result is accurate, efficient information processing in less time.

3.2 Invoice Imaging

It is widely used in many businesses applications to keep track of financial records and prevent a backlog of payments [100] from piling up. In government agencies and independent organizations, OCR simplifies data collection and analysis, among other processes. As the technology continues to develop, more and more applications are found for OCR technology, including increased use of handwriting [101] recognition. Furthermore, other technologies related to OCR, such as barcode recognition, are used daily in retail and other industries.

3.3 Legal Industry

In the legal industry, there has also been a significant movement to digitize paper documents. In order to save space and eliminate the need to sift through boxes of paper files, documents are being scanned and entered into computer databases. OCR further simplifies the process by making documents text-searchable, so that they are easier to locate and work with once in the database. Legal professionals now have fast, easy access to a huge library of documents in electronic format, which they can find simply by typing in a few keywords [100] [101].

3.4 Banking

The uses of OCR vary across different fields. One widely known application is in banking, where OCR is used to process checks without human involvement. A check can be inserted into a machine, the writing on it is scanned instantly, and the correct amount of money is transferred. This technology [100] [101] has nearly been perfected for printed checks, and is fairly accurate for handwritten checks as well, though it occasionally requires manual confirmation. Overall, this reduces wait times in many banks.

3.5 Healthcare

It Healthcare has also seen an increase in the use of OCR technology to process paperwork. Healthcare professionals always have to deal with large volumes of forms for each patient, including insurance forms as well as general health forms. To keep up with all of this information, it is useful to input relevant data into an electronic database that can be accessed as necessary. Form processing [100] [101] tools, powered by OCR, are able to extract information from forms and put it into databases, so that every patient's data is promptly recorded. As a result, healthcare providers can focus on delivering the best possible service to every patient.

3.6 Captcha

CAPTCHA is a program that can generate and grade tests that human can pass but current computers programmers' cannot. Hacking is a serious threat to internet usage. Now a day's most of the human activities like economic transactions, admission for education, registrations, travel bookings etc are carried out through

internet and all this requires a password which is misused by hackers. They create programs to like dictionary attacks and automatic false enrolments which lead to waste of memory and resources of website. Dictionary attack is attack against password authenticated systems where a hacker [102] writes a program to repeatedly try different passwords like from a dictionary of most common passwords. In CAPTCHA, an image consisting of series of letters of number is generated which is obscured by image distortion techniques, size and font variation, distracting backgrounds, random segments, highlights and noise in the image. This system can be used to remove this noise and segment the image to make the image tractable for the OCR (Optical Character Recognition) systems.

3.7 Institutional Repositories and Digital Libraries

Institutional repositories are digital collections of the outputs created within a university or research institution. It is an online locale of intellectual data of an institution, especially a research institution where it is collected, preserved and aired. It helps to open up the outputs of an institution and give it visibility and more impact on worldwide level. Enables and encourages interdisciplinary approaches to research and facilitates the development [103] and sharing of digital teaching materials and aids. It is basically a collection of peer reviewed journal articles, conference proceedings, research data, monographs, books, theses and dissertations and presentations. Their first role is to provide the Open Access literature. Practical implementation of this includes setting up a system which consists of scanner which scans the documents. This scanned document is then fed as an input to an Optical Character Recognition system where information is acquired and retained in digitized form.

3.8 Optical Music Recognition

Automated learning system extract information from images and is part of major researches. Optical music recognition (OMR) born in 1950's is a developed field and initially was aimed towards recognizing printed sheets [104] which can be edited into playable form with the help of electronic and electrochemical methods. An OMR system has many applications like processing of different classes of music, large scale digitization of musical data and also it can be used for diversity in musical notation. Image enhancement and segmentation is the basic step and hence the paper focuses on it.

3.9 Automatic Number Recognition

Automatic number plate recognition is used as a mass surveillance technique making use of optical character recognition on images to identify vehicle registration plates. ANPR has also been made to store the images captured by the cameras including the numbers captured from license [105] plate. ANPR technology own to plate variation from place to place as it is a region specific technology. They are used by various police forces and as a method of electronic toll collection on pay-per-use roads and cataloging the movements of traffic or individuals.

3.10 Handwriting Recognition

Handwriting recognition is the ability of a computer to receive and interpret intelligible handwritten input from sources such as paper documents, photographs, touch-screens and other devices. The image of the written text may be sensed "off line" from a piece of paper by optical scanning (optical character recognition) or intelligent word recognition. Alternatively, the movements of the pen tip may be sensed "on line", for example by a pen-based computer screen surface [106].

3.11 Other Industries

OCR is widely used in many other fields, including education, finance, and government agencies. OCR has made countless texts available online, saving money for students and allowing knowledge to be shared. Invoice imaging applications are used in many businesses to keep track of financial records and prevent a backlog of payments from piling up. In government agencies and independent organizations, OCR simplifies data collection and analysis, among other processes. As the technology continues to develop, more and more applications are found for OCR technology [100], including increased use of handwriting recognition. Furthermore, other technologies related to OCR, such as barcode recognition, are used daily in retail and other industries. To learn more about OCR solutions for your office, you can download a free trial of Maestro Recognition Server, CVISION's OCR toolkit, or Trapeze, our automated form-processing solution.

IV CONCLUSION

Nowadays, a lot of documents are produced in paper form but it is obvious, that automatic data recognition systems are very popular. The document is repeatedly copied and changed during subsequent processing steps, so it exists in many different copies. In some applications they can successfully help humans, but in some cases they are useless. Though researchers have suggested various sophisticated ideas and techniques to deal with the recognition of unconstrained and connected characters, practical OCR systems suffer from a lack of such characteristics. It is because of the claims made by the researchers are not adequately substantiated by exposure of the systems into real working environments/conditions the lack of practical feasibility of such advanced techniques with the available hardware from an economical viewpoint. From these constraints and the lack of performances it can be concluded that the ability to read text by machines with the same fluency as the human remains an unachieved goal, though a great amount of effort has already been expended on the subject. By handwritten character recognition one means the recognition of single and unconstrained hand drawn characters, i.e. numerals, upper-case and lowercase characters of a particular alphabet. However, the frontiers of character recognition have now moved to the recognition of cursive script that is the recognition of characters which may be connected or written in calligraphy.

REFERENCES

[1] Amarjot Singh, Ketan Bacchuwar, Akshay Bhasin, “A Survey of OCR Applications”, International Journal of Machine Learning and Computing, Vol. 2, No. 3, June 2012.

[2] Tappert, C. C. Suen, C. Y.; Wakahara, T., "The state of the art in online handwriting recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(8): 787.

[3] V. K. Govindan, A. P. Shivaprasad, "Character Recognition – A Review", Pattern Recognition, Vol. 23. No. 7, pp. 671-683 (1990).

[4] K. Bachuwar, A. Singh, G. Bansa, S. Tiwari, "An Experimental Evaluation of Preprocessing Parameters for GA Based OCR Segmentation" in 3rd International Conference on Computational Intelligence and Industrial Applications (PACIIA 2010), 2010, proceedings, Vol. 2, pp. 417 -420.

[5] M.D. Ganis, C.L. Wilson, J.L. Blue, "Neural network-based systems for handprint OCR applications" in IEEE Transactions on Image Processing, 1998, Vol: 7, Issue: 8, p.p. 1097 - 1112.

[6] R. Gossweiler, M. Kamvar, S. Baluja, "What's Up CAPTCHA? A CAPTCHA Based On Image Orientation", in WWW, 2009.

[7] B. Joanna," Building an institutional repository at Loughborough University: some experiences, program: Electronic library and information systems, 2009.

[8] A. Singh, K. Bacchuwar, A. Choubey, S. Karanam, D. Kumar, "An OMR Based Automatic Music Player", in 3rd International Conference on Computer Research and Development (ICCRD 2011) in, (IEEE Xplore), 2011, Vol. 1, pp. 174-178.

[9] S.L. Chang, T. Taiwan, L.S. Chen, Y.C. Chung, S.W. Chen, "Automatic license plate recognition" in IEEE Transactions on Intelligent Transportation Systems, 2004, Vol: 5, Issue: 1, p.p. 42 – 53.

[10] R Plamondon, S. N. Srihari, "On-line and off-line handwriting recognition: a comprehensive survey" IEEE transaction on pattern Analysis and machine Intelligence, 2000, 22(1), 63-84.

[11] Schantz, Herbert F., "The history of OCR, optical character recognition", [Manchester Center, Vt.]: Recognition Technologies Users Association (1982).

[12] d'Albe, E. E. F., "On a Type-Reading Optophone", Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 90 (619): 373–375, July 1914.

[13] M. E. Stevens, "Introduction to the special issue on optical character recognition (OCR)", Pattern Recognition, 2, 147-150 (1970).

[14] J. C. Rabinow, "Whither OCR and whence?", Datamation, 38-42, July (1969).

[15] P. L. Andersson, "Optical Character Recognition—A Survey", Datamation, 43-48, July (1969).

[16] P. L. Andersson, "OCR enters the practical stage", Datamation, 22-27, Dec. (1971).

[17] Magnetic ink character recognition system with angled read head, IBM Technical Disclosure Bull. 28, 2555- 2556 (1985).

[18] J. Mantas, "An overview of character recognition methodologies", Pattern Recognition 19, 425-430 (1986).

[19] Character Recognition. British Computer Society, Lon- don, England (1971).

[20] K. S. Fu, "Syntactic Pattern Recognition and Applications", Prentice Hall, Engiewood Cliffs, New Jersey (1982).

[21] G. Nagy, "Optical character recognition: theory and practice", Handbooks of Statistics, P. R. Kilshnaiah and L. N. Kanal, Eds, Vol. 2, pp. 621-649 (1982).

[22] S. N. Srihari, "Computer Text Recognition and Error Correction", IEEE Computer Society Press, Silver Spring, MD (1984).

[23] J. R. Ulimann, "Advances in character recognition", "Application of Pattern Recognition", K. S. Fu, Ed., pp. 197-236. CRC Press, Boca Raton, FL (1982).

[24] V. A. Kovalevsky, Ed., "Character Readers and Pattern Recognition", Spartan Books, New York (1968).

[25] Y. A. Kovalevsky, "Image Pattern Recognition", Springer, Berlin (1977).

[26] Optical Character Recognition and the Years Ahead. The Business Press, Illinois, U.S.A. (1969).

[27] C. Y. Suen and R. D. Moil, Eds, "Computer Analysis and Perception", Vol. 1: Visual Signals. CRC Press, Boca Raton, FL (1982).

[28] G. Nagy, "State of the art in pattern recognition", Proc. IEEE 56, 836-860 (1968).

[29] L. D. Harmon, "Automatic recognition of print and script", Proc. IEEE 60, 1165-1176 (1972).

[30] W. Stallings, "Approaches to Chinese character recognition", Pattern Recognition 8, 87-98 (1976).

[31] C. Y. Suen, M. Berthod and S. Mori, "Automatic recognition of handprinted characters--the state of the art", Proc. IEEE 68, 469-485 (1980).

[32] S. Moil, K. Yamamoto and M. Yasuda, "Research on machine recognition of handprinted characters", IEEE Trans. Pattern. Anal. Mach. Intell. 6, 386-405 (1984).

[33] R. H. Davis and J. L. Yall, "Recognition of handwritten characters--a review", image Vision Comput. 4. 208- 218 (1986).

[34] B. N. Chatterji, "Feature extraction methods for character recognition", IETE Tech. Rev. 3, 9-22 (1986).

[35] R. L. Grimsdale, F. H. Sumner, C. J. Tunis and T. Kilburn, "A system for the automatic recognition of patterns", Proc. I EE 106, 210-221 (1959).

[36] M. Eden, "On the formalization of handwriting, Structure of Language and its Mathematical Aspects", pp. 83-88, American Academic Society (1961).

[37] M. Eden, "Handwriting generalization and recognition, Recognizing Patterns, Kolers and M. Eden, Eds, pp. 138-154, M.I.T. Press, Cambridge, MA (1968).

[38] B. Blesser, R. Shillman, T. Kuklinski, C. Cox, M. Eden and J. Ventura, "A theoretical approach for character recognition based on phenomenological attributes", Proc. 1st Int. J. Conf. Pattern Recognition, Washington, pp. 33-40 (1973).

[39] C. Cox, B. Blesser and M. Eden, "The application of type font analysis to automatic character recognition", Proc. 2nd Int. J. Conf. Pattern Recognition, Copenhagen, pp. 226-232 (1974).

[40] R. J. Shillman, T. T. Kuklinski and B. A. Blesser, "Experimental methodologies for character recognition based on phenomenological attributes", Proc. 2nd Int. J. Conf. Pattern Recognition, Copenhagen, pp. 195-201 (1974).

[41] M. Yoshida and M. Eden, "Handwritten Chinese character recognition by an analysis by synthesis method", Proc. 1st Int. J. Conf. Pattern Recognition, Washington, pp. 197-204 (1973).

[42] M. Berthod, "Online analysis of cursive writing", Computer Analysis and Perception, Vol. 1, Visual Signals, C.Y. Suen and R. D. Mori, Eds. CRC Press, Cleveland, Ohio, U.S.A. (1982).

[43] R. Narasimhan, "Labelling schemata and syntactic description of pictures", Inform. Contr. 7, 151-179 (1964).

[44] R. Narasimhan, "Syntax directed interpretation of class of pictures", Commun. ACM 9, 166-173 (1966).

[45] R. Narasimhan, "On the description, generation and recognition of classes of pictures", Automatic Interpretation and Classification of Images, A. Gasselli, Ed., pp. 1-42. Academic Press, New York (1969).

[46] R. Narasimhan and V. S. N. Reddy, "A syntax-aided recognition scheme for handprinted English letters", Pattern Recognition 3, 345-361 (1971).

[47] T. Pavlidis and F. Ali, "Computer recognition of handprinted numerals by polygonal approximation", IEEE T. Syst. Man. Cyb. 5, 610-614 (1975).

[48] F. Ali and T. Pavlidis, "Syntactic recognition of handwritten numerals", IEEE T. Syst. Man Cyb. 7, 537-541 (1977).

[49] T. Pavlidis and S. L. Horowitz, "Segmentation of plaecurves", IEEE T. Comput. 23, 860-870 (1974).

[50] H. F. Feng and T. Pavlidis, "Decomposition of polygons into simpler components: Feature generation for syntactic pattern recognition, IEEE T. Comput. 24, 636-650 (1975).

[51] R. Casey and G. Nagy, "Recognition of printed Chinese characters", IEEE T. Elec. Comput. 15, 91-101 (1966).

[52] T. Agui and N. Nagahashi, "A description method of handprinted Chinese characters", IEEE T. Pat. Anal. Mach. Intell. 1, 20-24 (1979).

[53] A. Arakawa, "Online recognition of handwritten characters - Alphanumerics and Hirakana, Katakana, Kanji, Pattern Recognition 16, 9-16 (1983).

[54] C. H. Leung, Y. S. Cheung and Y. L. Wong, "A knowledge based stroke-matching method for Chinese characters", IEEE T. Syst. Man Cyb. 17, 993-1003 (1987).

[55] Y. Yong, "Handprinted Chinese character recognition via neural networks", Pattern Recognition Lett. 7, 19-25 (1988).

[56] I. K. Sethi and B. Chatterjee, "Machine recognition of handprinted Devanagari numerals", J. Inst. Elec. Telecom. Engng (India) 22, 532-535 (1976).

[57] I. K. Sethi, Machine recognition of constrained handprinted Devanagari", Pattern Recognition 9, 69-75 (1977).

[58] R. M. K. Sinha and H. Mahabala, Machine recognition of Devanagari script, IEEE T. Syst. Man Cyb. 9, 435-449 (1979).

[59] R. M. K. Sinha, "Role of context in Devanagari script recognition", J. Inst. Elec. Telecom. Engng (Indian) 33, 86-91 (1987).

[60] R. M. Sinha, "Role of contextual postprocessing for Devanagari text recognition", Pattern Recognition 20, 475-485 (1987).

[61] G. Siromoney, R. Chandrasekaran and M. Chandrasekaran, "Machine recognition of printed Tamil characters", Pattern Recognition 10, 243-247 (1978).

[62] G. Siromoney, R. Chandrasekaran and M. Chandrasekaran, Machine recognition of Brahmi script, IEEE T Syst. Man Cyb. 13, (1983).

[63] M. Chandrasekaran, R. Chandrasekaran and G. Siromoney, "Context dependent recognition of handprinted Tamil characters", Proc. Int. Conf. Syst. Man Cyb., (India) 2, 786-790 (1984).

[64] R. Chandrasekaran, M. Chandrasekaran and G. Siromoney, "Computer recognition of Tamil, Malayalam and Devanagari characters", J. Inst. Elec. Telecom. Engng (India) 30, 150-154 (1984).

[65] P. Chinnuswamy and S. G. Krishnamoorthy, "Recognition of handprinted Tamil characters", Pattern Recognition 12, 141-152 (1980).

[66] S. N. S. Rajasekaran and B. L. Deekshatulu, "Recognition of printed Telugu characters", Comput. Graph. Image Process. 6, 335-360 (1977).

[67] A. K. Ray and B. Chatterjee, "Design of a nearest neighbour classifier system for Bengali character recognition", J. Inst. Elec. Telecom. Engng (India) 30, 226-229 (1984).

[68] H. Tanaka, Y. Hirakawa and S. Kaneku, "Recognition of distorted patterns using Viterbi algorithm", IEEE T. Pattern Anal. Mach. Intell. 4, 18-25 (1982).

[69] G. P. R. Sarvarayudu and I. K. Sethi, "Walsh descriptors for polygonal curves", Pattern Recognition 16, 327-336 (1983).

[70] M. Shridhar and A. Badreldin, High accuracy character recognition algorithm using Fourier and topological descriptors, Pattern Recognition 17, 515-523 (1984).

[71] M. Shridhar and A. Badreldin, "A high accuracy syntactic recognition algorithm for handwritten numerals", IEEE T. Syst. Man Cyb. 15 (1985).

[72] K. Sato, I. Isshiki, A. Ohoka and K. Yoshida, "Hand-scan OCR with a one-dimensional image sensor", Pattern Recognition 16, 459-467 (1983).

[73] C. J. Evangelisti, "Some experiments in the evaluation of character recognition scanners", Pattern Recognition 16, 273-287 (1983).

[74] G. Nagy, S. Sethi and Einspahr, "Decoding substitution ciphers by means of word matching with application to OCR", IEEE T. Pattern Anal. Mach. Intell. 9, 710-715 (1987).

[75] H. Almuallim and S. Yamaguchi, A method of recognition of Arabic cursive handwriting, IEEE T. Pattern Anal. Mach. Intell. 9, 715-722 (1987).

[76] T. S. El-Sheikh and R. M. Guindi, "Computer recognition of Arabic cursive scripts, Pattern Recognition 21, 293-302 (1988).

[77] J. J. Hull, "Word shape analysis in a knowledge based system for reading text", 2nd Conf. on Artificial Intelligence Applications: The Engineering of Knowledge Based System, Miami Beach, FL, U.S.A., pp. 114-119 (11-13 December 1985).

[78] K. Aoki and Y. Yamaya, "Recognizer with learning mechanism for handwritten English script words", Proc. 8th Int. J. Conf. Pattern Recognition, Paris, France, pp. 690-692 (27-31 October 1986).

[79] K. H. Wong and F. Fallside, "Dynamic programming in the recognition of connected handwritten scripts", 2nd Conf. Artificial Intelligence Applications: The Engineering of Knowledge Based Systems, Miami Beach, FL, U.S.A., pp. 666-670 (11-13 December 1985).

[80] S. N. Srihari and R. M. Bozinovic, "A multilevel perception approach to reading cursive script", Artificial Intell. (Netherlands) 33, 217-255 (October 1987).

[81] K. R. Tampi and S. S. Chetlur, Segmentation of handwritten characters, Proc. 8th Int. J. Conf. Pattern Recognition, Paris, France, pp. 684-686 (27-31 October 1986).

[82] V. R. Ting and R. K. Ward, "Separation and recognition of connected handprinted English characters", IEEE Pacific Rim Conf. Communications, Computers and Signal Process. Proc., Victoria, BC, Canada, 4-5 June 1987, pp. 512-516 (1987).

[83] L. Lam and C. Y. Suen, "Structural classification and relaxation matching of totally unconstrained handwritten ZIP-code numbers", Pattern Recognition 21, 19-31 (1988).

[84] G. Baptista and K. M. Kulkarni, "A high accuracy algorithm for recognition of handwritten numerals", Pattern Recognition 21, 287-291 (1988).

[85] J. J. Cannat and Y. Kodratoff, "Learning technique applied to multifont character recognition", Proc. SPIE, Int. Soc. Opt. Engng 635, 469-479 (1986).

[86] J. J. Cannat, Y. Kodratoff and S. Moscatelli, "Learning techniques applied to multifont character recognition", Proc. 8th Int. J. Conf. Pattern Recognition, pp. 123- 125 (1986).

[87] R. Malyan and R. Sunthankar, "Handprinted text reader that learn by experience", Microprocessor Microsystem (U.K.) 10, 377-385 (1986).

[88] C. G. Leedham and A. C. Downton, "On-line recognition of Pitman's handwritten shorthand -an evaluation of potential", Int. J. Man Mach. Stud. (U.K.) 24, 375-393 (1986).

[89] C. G. Leedham and A. C. Downton, "Automatic recognition and transcription of Pitman's handwriting shorthand-an approach to short forms", Pattern Recognition 20, 341-348 (1987).

[90] H. Nagahashi and M. Nakatsuyama, "A pattern description and generation method for structural characters", IEEE T. Pattern Anal. Mach. Intell. 8, 112-118 (1986).

[91] Harjinder Singh, Description aided recognition of handprinted characters, Ph.D. Thesis, Indian Institute of Science, Bangalore, India (1985).

[92] J. S. Huang and M. Lung, Separating similar complex Chinese characters by Walsh transform, Pattern Recognition 20, 425-428 (1987).

[93] G. Wolberg, "A syntactic Omni-font character recognition system", Proc. CVPR '86: IEEE Comput. Society Conf on Computer Vision and Pattern Recognition, Miami Beach, FL, U.S.A,, pp. 168-173 (22-26 June 1986).

[94] S. W. Lam and H. S. Baird, "Performance testing of mixed font variable size character recognizers", Proc. 5th Scandinavian Conf. Image Analysis, Stockholm, Sweden, Vol. 2, pp. 563-570 (2-5 June 1987).

[95] T. Pavlidis, "A vectorizer and feature extractor for document recognition", Comput. Vision Graph. Image Process. 35, 111-127 (1986).

[96] H. Hagita and I. Masuda, "Design principles of feature vectors for recognition of large character sets", Proc. 1987 Int. Conf. Syst. Man Cyb., Alexandria, VA, U.S.A., Vol. 2, pp. 826-830 (20-23 October 1987).

[97] V. K. Govindan, "Computer Recognition of Handprinted Characters: An Automated Approach to the Design of Recognizers", Ph.D. Thesis, Dept of Electrical Communication Engineering, Indian Institute of Science, Bangalore, India (1988).

[98] H. Kami, "Evaluation of automatic dictionary generation for character recognition", NEC Res. Demi. pp. 42-47 (January 1984).

[99] K. Bachuwar, A. Singh, G. Bansa, S. Tiwari, "An Experimental Evaluation of Preprocessing Parameters for GA Based OCR Segmentation", Proceedings of 3rd International Conference on Computational Intelligence and Industrial Applications (PACIIA 2010), Vol. 2, pp. 417 -420, 2010.

[100] http://www.cvisiontech.com/reference/general-information/ocr-applications.html.

[101] M.D. Ganis, C.L. Wilson, J.L. Blue, "Neural network-based systems for handprint OCR applications", IEEE Transactions on Image Processing, Vol: 7, Issue: 8, p.p. 1097 – 1112, 1998.

[102] R. Gossweiler, M. Kamvar, S. Baluja, "What's Up CAPTCHA? A CAPTCHA Based On Image Orientation", WWW, 2009.

[103] B. Joanna, "Building an institutional repository at Loughborough University: some experiences, program:", Electronic library and information systems, 2009.

[104] A. Singh, K. Bacchuwar, A. Choubey, S. Karanam, D. Kumar, "An OMR Based Automatic Music Player", 3rd International Conference on Computer Research and Development (ICCRD 2011), (IEEE Xplore), Vol. 1, pp. 174-178, 2011.

[105] S.L. Chang, T. Taiwan , L.S. Chen, Y.C. Chung, S.W. Chen, "Automatic license plate recognition", IEEE Transactions on Intelligent Transportation Systems, Vol: 5 , Issue: 1, p.p. 42 – 53, 2004.

[106] R Plamondon, S. N. Srihari, "On-line and off-line handwriting recognition: a comprehensive survey", IEEE transaction on pattern Analysis and machine Intelligence, 22(1), 63-84, 2000.