IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 2, April 2014.
www.ijiset.com
ISSN 2348 - 7968

Automatic Character Recognition of Indian Languages – A brief Survey

Ms.M.Shalini[1], Dr.B.Indira[2]
Kasturba Gandhi Degree & PG college for Women,
Secunderabad, Andhra Pradesh, India

Abstract: Optical Character Recognition (OCR) is a very important task in Pattern Recognition. Foreign languages, especially English character recognition has been extensively studied by many researches but in case of Indian Languages which are complicated, the research work is very limited and constrained. This paper presents a brief survey of earlier research work related to all Indian languages. A brief history of OCR, various approaches to character recognition along with some applications of character recognition is also discussed in this paper.

Keywords: Pattern Recognition, Optical Character Recognition, Template matching, Artificial Neural Networks.

Introduction

Optical Character Recognition is an active field of research in Pattern Recognition. The problem of character recognition can be classified based on two criteria. One is based on the type of the text which is printed or hand written. The other is based on the acquisition process which can be on-line of off-line [1]. It is generally considered that the on-line method of recognizing handwritten text has achieved better results than its off-line counterpart. This may be attributed to the fact that more information may be captured in the online case such as the direction, speed and the order of strokes of the handwriting. On the other hand machine-printed character recognition can achieve very good results on good quality documents.

In case of online character recognition, there is real time recognition of characters. Online systems have better information for doing recognition since they have timing information and can avoid the initial search step of locating the character as in the case of their offline counterpart. Online systems obtain the position of the pen or printed character as a function of time directly from the interface. Offline recognition of characters is known as a challenging problem because of the complex character shapes and great variation of character symbols written or printed in different modes.

In case of offline character recognition, the typewritten or handwritten character is typically scanned in the form of a paper document and made available in the form of a binary or gray scale image to the recognition algorithm. Offline character recognition is a more challenging and difficult task as there is no control over the medium and instrument used. Subsequent operations such as scanning and binarization present additional challenges to the algorithm for the offline character recognition. The major difference between Online and Offline Character Recognition is that Online Character Recognition has real time contextual information but offline data does not. This difference generates a significant divergence in processing architectures and methods.

Brief History of Character Recognition

Many methods have been proposed for character recognition. But they are often subjected to substantial constraints due to unexpected difficulties. Historically character recognition system has evolved in three ages [2], namely the periods cited denoting as

1900-1980 (early ages) – The history of character recognition can be traced as early as 1900. When the Russian Scientist Tyering attempted to develop an aid for visually handicapped. The first character recognizers appeared in the middle of 1940s with the development of digital computers. The early work on the automatic recognition of characters has been concentrated either upon machine printed text or upon small set of well distinguished hand written text or symbols. The commercial character recognizers were available in 1950s.

1980-1990 Developments – The studies until 1980 suffered from the tack of powerful computer hardware and data acquisition derives. However, the character recognition research was focused on basically the shape recognition techniques without using any semantic information.

After 1990 advancements – The real progress on character recognition system is achieved during this period, using the new development tools and methodologies, which are empowered by continuously growing information technologies. In the early nineties, Image processing and Pattern recognition techniques are efficiently combined with the Artificial Intelligence methodologies. Nowadays in addition to the more powerful computers and more accurate electronic equipments such as scanners, cameras and electronic tablets, we have efficient, modern use of methodologies such as neural

ISSN 2348 - 7968

networks, Hidden Markov models; Fuzzy set reasoning and Natural language processing.

Character recognition system is the base for many different types of applications in various fields, many of which we use in our daily lives. Post offices, banks, security systems, number plate recognition system and even in the field of robotics use this system as the base of their operations.

Character Recognition Approaches

Character recognition systems extensively use the methodologies of pattern recognition, which assigns an unknown sample to a predefined class. Many techniques for character recognition are investigated by the researchers and character recognition approaches can be classified as [3] Template matching, Statistical techniques, Syntactic or structural, Neural network, Hybrid or Combination approaches.

Template matching approach

This is the simplest way of character recognition, based on matching the stored data against the character to be recognized. The matching operation determines the degree of similarity between two vectors i.e. group of pixels, shapes curvature etc. a gray level or binary input character is compared to a standard set of stored data set. According to similarity measure (e.g. Euclidean, Yule similarity measures etc.), a template matcher can combine multiple information sources, including match strength and k-nearest neighbor measurements from different matrices. The recognition rate of this method is very sensitive to noise and image deformation.

Statistical Techniques

Statistical decision theory is concerned with statistical decision functions and a set of optimality criteria, which maximizes the probability of the observed pattern given the model of a certain class. Statistical techniques are based on the assumptions such as Distribution of the feature set, statistics available for each class, collection of images to extract a set of features which represents each distinct class of patterns. The measurements taken from n-features of each word unit can be thought to represent an n-dimensional vector space. The major statistical methods applied in the character recognition field are Nearest Neighbor Likelihood or Bayes classifier, clustering Analysis, Hidden Markov Modeling, Fuzzy Set Reasoning, Quadratic classifier etc.

Syntactic or Structural Approach

In Syntactic Pattern recognition a formal analogy is drawn between the structure of pattern and syntax of a language. Structural pattern recognition is intuitively appealing because in addition to classification, this approach also provides a description of how the given path constructed from the primitives. Flexible structural matching is proposed for identification of alphanumeric characters.

Neural Networks

Various types of neural networks are used for character recognition classification. A neural network is a computing architecture that consists of massively parallel interconnection of adaptive neural processors. Because of its parallel nature, it can perform computations at a higher rate compared to classical techniques. Because of its adaptive nature, it can adapt to changes in the data and learn the characteristics of input signal. Output from one node is fed to another one in the network and final decision depends on the complex interaction of all nodes. Several approaches exist for training of neural networks like error correction, Boltzman, Hebbian and competitive learning. Neural network architectures can be classified as, feed-forward, feed-back and recurrent networks. The most common neural networks used in the character recognition systems are the Multi Layer Perceptron (MLP) of the feed forward networks and the Kohonen's Self Organizing Map of the feedback networks.

Hybrid or Combination Classifier

We may have different feature sets, different training sets, different classification methods or different training sections, all resulting in set of classifiers, whose outputs may be combined, with the hope of improving overall classification accuracy. If this set of classifiers is fixed, the problem focuses on the combination function. It is also possible to use a fixed combiner and optimize the set of input classifiers. A typical combination scheme consists of a set of individual classifiers and combiner which combines the results of the individual classifiers to make the final decision. Various schemes for combining multiple classifiers can be grouped into three main categories according to their architecture parallel, cascading, and hierarchical.

Foreign Languages

A large amount of research work has been carried out on Chinese character recognition. Chinese character recognition is considered to be a very hard problem and regarded as one of the ultimate goals of character recognition research. The various techniques employed in the recognition of Chinese characters can be found in the review work [4]. Broadly the techniques employed are: statistical, structural, syntactical and neural networks [5-11].

Mu-Chun Su et.al. [12], proposed a neural network based approach to optical symbol recognition. They demonstrated that node heads could be easily recognized by using a set of fuzzy

ISSN 2348 - 7968

rules extracted from the parameters of trained neural networks and also showed that only12 features are sufficient to achieve a high recognition rate.

In [13], the authors Nada et.al., presented Quick Stroke which is a system for the incremental recognition of handwritten Chinese characters. Incremental recognition is an approach for on-line recognition of ideographic characters. It allows a user to enter characters a factor of 2 times faster than systems that require entry of full characters. Incremental recognition is performed by a two-stage system which utilizes 68 neural networks with more than 5 million free parameters. To enable incremental recognition, a time-delay neural network (TDNNs) is used that are trained to recognize partial characters. Quick Stroke is 97.3% accurate for the incremental writer-independent recognition of 4400 simplified GB Chinese ideograms.

Arabic, like Hebrew is written from right to left. Arabic text is cursive in general i.e. Arabic letters are normally connected on the base line. Therefore, the recognition rate of Arabic characters is lower than that of disconnected characters such as printed English. There are many research works on the recognition of printed and hand written Arabic characters [14-21].

An automatic off-line character recognition system [22] for handwritten cursive Arabic characters is proposed by Abuhaiba et.al.,. The developed algorithm yields skeletons that reflect the structural relationships of the character components. The character skeleton is converted to a tree structure suitable for recognition. A set of fuzzy constrained character graph models (FCCGM's) is designed. These models are graphs, with fuzzily labeled arcs used as prototypes for the characters. A set of rules was applied in sequence to match a character tree to an FCCGM.

An invariant mapping is applied to hand-written Arabic character recognition by Nawwaf et.al.,[23]. In this paper an application of mapping is used which produces the same output pattern regardless of the orientation, position and size of the input pattern. The mapping has the advantage of being simple.

Fiaz Hussain and John Cowell [24], developed a system for Character Recognition of Arabic and Latin Scripts. They also discussed the main components used in the multi-stage system, paying particular attention to the normalization process used for orientation and size for a given bitmapped character.

Cherry Blossom, a system for reading unconstrained handwritten page images by Sri hari et.al.,[25], is a Japanese OCR system developed at CEDAR. It is a full-featured prototype system that can process a scanned document page and output recognized text in JIS code. The system is designed to process a large variety of documents with degraded print quality. The lowest allowable scan resolution is 200 dpi. The character classifier can recognize 3,377 Japanese characters which includes the first level Kanji, Hiragana, Katakana, alpha numerals and other symbols. Related works in Japanese character recognition are reported in [26-28].

Some research work, even though not exhaustive, pertaining to the Korean and Thai characters are presented in [29-31].

Indian Character Recognition

Not many attempts have been made on the character recognition of Indian character sets. However, some major works are reported on Devanagari. Some attempts are also reported on Tamil, Kannada, Gujarathi, Bengali, Malayalam and Telugu.

Character recognition of handwritten and printed text is of great importance for electronic conversion of historical information including letters, diaries, wills and other manuscripts. The problem is challenging because of human handwriting variability, uneven skew and orientation as well as noise and distortion such as smudges, smears, faded print, etc. identification of handwritten Indian scripts especially of Bangla, as well as English, Hindi, Malayalam, etc. Most of the Indian scripts have 500 or more characters or symbols used in running text, through the number of basic vowels and consonants is not more than 50. The number is multiplied by three types of vowel modifiers that may be glued below the consonants, thus generating threefold consonant-vowel combinations. Further increase in number is possible where consonant creates a complex orthographic shape called compound characters. For some scripts like Bangla, Gujarthi, Telugu and Devanagari languages consists of large number of compound characters. These compound characters can also take vowel modifiers to generate threefold more shapes. Thus orthographic shapes may run of the order of thousand. Only Tamil and Punjabi scripts are relatively simpler, where the number of characters/ symbol is about 150 and 70 respectively. Most Indian script lines can be partitioned into three sub-zones. The upper and lower zones may consist of parts of the basic characters as well as vowel modifiers. These parts of two consecutive text lines normally do not overlap or touch in case of printed script, but for handwriting, people have the tendency to write them bigger, leading to overlapping and

touching characters. Overall these characteristics make handwritten and printed Indian text recognition more challenging.

The earlier research work on character recognition related to Indian languages is discussed below.

- **Telugu Character Recognition**

A two stage recognition method for printed Telugu characters is proposed by Rajasekaran and Deekshatulu [32] . In the first stage, the primitive shapes are removed from the given Telugu character using a syntax-aided recognition scheme. After removal of the primitives only basic shapes were left. In the second stage, the basic letters can be recognized using either the classificatory or syntactic approach. The basic letters are recognized by a process called On the Curve Coding. The classification is achieved by means of a decision tree by using the knowledge of the primitives and the basic characters that are present in the input pattern. The extraction of primitive technique, being a very complex one, takes more time for recognition processing.

In [33] Sukhaswamy et.al., proposed an approach to recognize Telugu script using Neural Networks. The author developed a network architecture called Multiple Neural Network Associative Memory (MNNAM) for recognition of Printed Telugu characters. In this method, the exemplars to be trained are divided into groups, each of which is having a capacity less than the practical optimal storage capacity (POSC) of a network. Each group is trained into a separate network of same topology. The test pattern to be recognized is presented to each of these networks. The patterns to which each of the networks converge are then made as exemplars to train further levels of networks, called as combination networks. The main limitation of their work is that the character recognition is not invariant of size, translation and rotation.

In [34], the authors P V S Rao and TM Ajitha have suggested Telugu script recognition using a feature based approach. Recognition is based on segmenting the characters into the component elements and identifying them. Feature vector parameters for individual basic characters are extracted from single specimen written in isolation. These are suitably combined to construct feature vectors for compound characters. These are compared with similar feature vectors extracted from the test samples to be recognized.

An OCR system for Telugu is developed by Atul Negi et.al.,[35]. They identified a total set of 370 connected components that exhaust each of the five categories, vowel, Consonants, vowel signs, conjunct consonants, vowel or consonant, combination of a consonant and vowel sign. Template matching is used to recognize the components. Frinze distance is used as a distance measure for comparison of images.

In [36] the authors presented a system to locate, extract and recognize Telugu text. First, the Hough Transform for circles is performed on the Sobel gradient magnitude of the image to locate text. The located circles are filled to yield text regions, followed by Recursive XY Cuts to segment the regions into paragraphs, lines and word regions. A region merging process with bottom-up approach envelopes individual words. Local binarization of the word MBRs yields connected components containing glyphs for recognition. The recognition process first identified candidate characters by a zoning technique and then constructs structural feature vectors by cavity analysis. Finally, if required, crossing count based non-linear normalization and scaling is performed before template matching. [37], the author proposed a method which uses wavelet multi resolution analysis for the purpose of extracting features and associative memory model to accomplish the recognition tasks. Wavelet Basis function is used to extract the invariant features of the characters and Hopfield-based Dynamic Neural Network model used for the purpose of learning and recognition.

Modular Neural Networks approach and minimum number of primitives in the training process for the recognition of complete set of printed Telugu characters is presented by Sandhya Rani[38]. Extraction of primitives, identifying the class of primitives and recognizing the characters using Neural Networks techniques are the significant works. Srivani [39], proposed a novel technique for processing Printed Telugu document using Modular Neural Network approach. The developed system is robust with rotation; size and noise efficient feature extraction techniques are adopted and achieved high recognition accuracy.

- **Tamil Character Recognition**

The Tamil alphabet consists of 12 vowels and 18 consonants. These combine to form 216 compound characters. There is one special character (*aaytha ezutthu*), giving a total of 247 haracters. Unlike other Indian languages, Tamil has single glyphs for ka, cha, ta, tha, pa, ra. But their sounds vary depending on the context where they occur. However, with the advantage of having a separate symbol for each vowel in composite character formations, there is a possibility to reduce the number of symbols used by the alphabet. In character

recognition point of view, only 67 symbols have to be identified to recognize all 247 characters. Some of the research works regarding Tamil character recognition are as follows.

Anbumani et.al., in [40] proposed an Optical Character Recognition of Printed Tamil Characters. Author used statistical parameters during recognition stage.

Ralph Niels and Louis Vuurpij [41] discussed the use of Dynamic Time Warping (DTW) for classifying handwritten Tamil characters. A prototype based classifier is proposed that uses DTW both for generating prototypes and for calculating a list of nearest prototypes. Prototypes were automatically generated and selected. Two tests were conducted to measure the performance of the classifier in a writer dependent and independent setting. Furthermore, several strategies were developed for rejecting uncertain cases. Two different rejection variables were implemented and using a Monte Carlo simulation, the performance of the system was tested in various configurations.

The author Fernando [42], proposed a system to recognize handwritten Tamil characters using a two-stage classification approach, for a subset of the Tamil alphabet. In the first stage, an unknown character is pre-classified into one of the three groups: core, ascending and descending characters. Then, in the second stage, members of the pre-classified group are further analyzed using a statistical classifier for final recognition. A recognition rate of 80% was achieved for the 1st choice and 97% for the top 3 choices. Related works in Tamil character recognition are presented [43-45].

- **Kannada Character Recognition**

The Kannada alphabet is classified into two main categories: vowels and consonants. There are 16 vowels and 35 consonants and words in Kannada are composed of *aksharas* which are analogous to characters in an English word. While vowels and consonants are *aksharas*, the vast majority of *aksharas* are composed of combinations of these in a manner similar to most other Indian scripts. A few research works on character recognition of kannada is as follows.

In [46], Kannada characters written on the digitizer pad are recognized in an interactive way is proposed by R. Srinivasa Rao and Sudhakar Samuel. A document file of recognized text in Kannada computer fonts is produced as one keeps writing the text on the digitizer pad. Wavelet transforms have been used to extract the features of the Kannada characters, which have turned to be more efficient features. The multilayer feed forward neural networks are used to recognize and classify the Kannada characters.

A font and size-independent OCR system for printed Kannada documents using support vector machines is presented by Ashwin and Sastry[47]. In this work recognition is achieved by employing a number of 2-class classifiers based on the Support Vector Machine (SVM) method.

- **Bengali Character Recognition :**

Bengali alpha numeric character recognition is proposed by Dutta et.al.,[48]. In this work, Curvature related characteristics were used as features and back propagation based learning scheme was used in the recognition strategy enables the system to learn from examples. Other works related to Bengali are reported in [49-51].

Few other research works are presented in other Indian languages like Gujarati, Gurumukhi [52, 53].

- **Hindi Character Recognition**

Devanagari script is alphabetic in nature and the words are two dimensional compositions of characters and symbols which makes it different from Roman and ideographic scripts.

I K Sethi [54] described Devanagari numeral recognition based on the structural approach. The primitives used are horizontal line segment, vertical line segment, right slant and left slant. A decision tree is employed to perform the analysis based on the presence/absence of these primitives and their interconnection. A similar strategy was applied to constrained hand printed Devanagari characters.

An OCR system for printed Devanagari script [55] is presented by Pal and Chaudhuri claims an accuracy of 95% at the character level. In this system some standard and some new techniques have been used for preprocessing. However, thinning has not been carried out on the images. From zonal information and shape characteristics, the basic, modified and compound characters are separated for convenience of classification. Modified and basic characters are recognized by a structural feature based binary tree classifier while the compound characters are aimed to be recognized by a hybrid approach.

In [56-58] the authors have carried out a few notable works in Devanagari script recognition. They presented a syntactic pattern analysis system with an embedded picture language for Devanagari script recognition. The system stores structural descriptions for each symbol of the script in terms of primitives and their relationships. The recognition involves a

search for the unknown character primitives based on the stored description. They also suggested knowledge based contextual post-processing systems for Devanagari text recognition.

Veena Bansali[59], designed a Devanagari text recognition system by integrating knowledge sources, features of characters such as horizontal zero crossings, moments, aspect ratios, pixel density in nine zones, number, and position of vertex points, with structural descriptions of characters. These were used to recognize the characters.

Huanfeng et.al., [60] presented a recognition system for Devenagari lipi based on the Generalized Hausdorff Image Comparison (GHIC), which is like a template matching method but over comes some disadvantages of the traditional template matching approach.

Classification and recognition of printed Hindi characters using Artificial Neural Networks is proposed by Indira et.al.,[61] and achieved a recognition rate in the range of 76-95% for various samples. The results also show that the recognition accuracy and efficiency of the network increases with more number of training samples.

Applications of Character Recognition System

Optical Character Recognition has a wide range of applications in various areas. It can be used as a telecommunication aid for postal address reading for the deaf, processing of documents, in recognition of foreign language and also for language translation [62]. In bill processing systems it is used to read payment slips like electricity bills , telephone / water bills. It will read and recognize the amount to be paid and also recognize the account number. The character recognition system can also be used for reading the address, assigning Zip codes to letters, application forms, voter ID cards, and identification of bank cheques by recognizing the account number and the amount written on the cheque. These systems can also be used in automatic processing of issuing tickets to air line passengers, validation of passports and visa cards etc. Address readers in postal departments locates the address on letters and sorts them according to their location using the zip code. The multiline optical character reader (MLOCR) by United States Postal Services (USPS) locates the address block on a mail piece, reads the address, identifies ZIP and generates a 9-digit bar code and sorts the mail to the correct stacker. This classifier recognizes up to 400 fonts and the system can process up to 45,000 mail pieces per hour [63].

Conclusion

Character Recognition is one of the vital tasks in Pattern Recognition. The popularity and use of Character Recognition is increasing day by day with the advent of new, fast and efficient hardware and software. But automatic character recognition of Indian languages is still in preliminary stage and hence there is a need of lot of research to address the various issues and their complexities. There are many factors such as noise, various font sizes, broken lines or characters, quality of the image, problems in segmentation that influence recognition process. India is a multi lingual country; so many more efficient and real-time text recognizers are required. A good text recognizer has many commercial and practical applications. Hence there is a need to develop a very good character recognition system which must achieve highest accuracy.

REFERENCES

[1] R. Plamondon and S. N. Srihari, “On-line and off-line handwritten recognition: a comprehensive survey”, IEEE Transactions on PAMI, Vol. 22(1), pp. 63–84, 2000.

[2] T. Yarman, Nafiz arica, factos – “An overview of CR focused on offline handwriting “– IEEE – 1996.

[3] K. Anil Jain, “Statistical Pattern Recognition: A Review”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 22, 1, 2000, pp. 4-37.

[4] W. Stallings, “Approaches to Chinese Character Recognition”, *Pattern Recognition*, Vol. 8, 1976, pp. 87-98.

[5] R. Casey and G. Nagy, “Recognition of Printed Chinese Characters”, *IEEE Trans. Elec. Compute*, 15, 1966, pp. 91-101.

[6] T. Agui and N. Nagahashi, “A Description Method of Hand Printed Chinese Characters”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 1979, pp. 20-24.

[7] T.H. Hildebrandt and W. Liu, "Optical Recognition of Handwritten Chinese Characters: Advances Since 1980", *Pattern Recognition*, Vol.26, 2, 1993, pp. 205-225.

[8] N. Fujui, H. Sugawara , E. Yamamoto, C. Ito and Fujita, “Some Results on Hand Printed Kanji Character Recognition Using the Features Extracted from Multiple Stand Point”, *IEEE Japan, PRL*,1981, pp.81-132.

[9] Y.S. Cheng and C.H. Leung, "Chain-Code Transform for Chinese
Character Recognition", *IEEE Proc. Int. Conf. Cyb. Soc*., Tucson, AZ, U.S.A., 1985, pp. 42-45.

[10] C.H. Leung, Y.S. Cheung and Y.L. Wong, "A Knowledge Based Stroke - Matching Method for Chinese Characters", *IEEE Trans. System Man and Cybernetics*, Vol. 17, 1987, pp. 993-1003.

[11] J.S. Huang and M.Lung, "Separating Similar Complex Chinese Characters by Walsh Transform", *Pattern Recognition*, Vol. 20, 1987, pp. 425-428.
[12] Mu-Chun Su, Hsin-Hua Chen, Wan-Chi Cheng, "A Neural Network Based Approach to Optical Symbol Recognition", *Neural Processing Letters 15*(2),pp: 117-135,2002.
[13] Nada P. Matic, John C. Platt, Tony Wang, "Quick Stroke: An Incremental On-line chinese Handwriting Recognition System", http:// research.microsoft.com/~jplatt/ ccr.pdf
[14] T.S. El. Sheikh and R.M. Guindi, "Computer Recognition of Arabic Scripts", *Pattern Recognition*, Vol. 21, 1968, pp. 293-302.
[15] B. Parhami and M. Taraghi, "Automatic Recognition of Printed Farsi texts", *Pattern Recognition*, Vol.14, 1981, pp. 395-403.
[16] A. Amin, "Machine Recognition of Handwritten Arabic Word by The IRAC II System", *6th Int. Conf. Pattern Recognition*, Munich, 1982, pp. 34-36.
[17] H. Almuallim and S. Yamaguchi, "A method of Recognition of Arabic Cursive Hand Writing", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 9, 1987, pp. 715-722. [18] S. Al-Emami and M. Usher, "On-line Recognition of Handwritten Arabic Characters", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12, 1990, pp. 704-710.
[19] H. Al-Yousefi and S.S. Udpa, "Recognition of Arabic Characters", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.14, 1992, pp.853-857.
[20] Adnan Amin, Humond AL Sadoun and Stephan Fischer, "Hand Printed Arabic Character Recognition System Using an Artificial Networks", *Pattern Recognition*, Vol. 29, 4, 1996, pp. 663-675.
[21] A. Amin and H.B. Al-Sadoun, "Hand Printed Arabic Character Recognition System", *12th Int. Conf. Pattern Recognition*, 1994, pp.536-539.

[22] S. I. Abuhaiba, S. A. Mahmoud, R. J. Green, "Recognition of Handwritten Cursive Arabic Characters", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol**.** 16(6), pp: 664 - 672 , 1994 .
[23] Nawwaf N. Kharma and Rabab K. Ward, "A Novel Invariant Mapping Applied to Handwritten Arabic Character Recognition", *Pattern Recognition*, Vol. 34, 11, 2001, pp. 2115-2120.
[24] Fiaz Hussain, John Cowell. "Character Recognition of Arabic and Latin Scripts," *iv*, p. 51, Fourth International Conference on Information Visualisation, 2000.
[25] S.N. Srihari, T. Hong and Z. Shi, "Cherry Blossom: A System for Reading Unconstrained Handwritten Page Images (PDF)", *Symposium on Document Image Understanding Technology (SDIUT)*, Annapolis, MD, April 1997.
[26] A. Arakawa, "Online Recongition of Handwritten Characters Alphanumeric and Hirakana, Katakana, Kanji", *Pattern Recognition*, Vol. 16, 1983, pp. 9-16.
[27] I. Sekita, K. Toraichi, R. Mori, K. Yamamoto and H. Yamada, "Feature Extraction of Hand Printed Japanese Characters by Spline Function for Relaxation Matching", *Pattern Recognition*, PP. 261-93, Vol. 21, 1988, pp. 9-17.
[28] Seii Chi Ozawa, Toshihide, Manabu Kotani, and Noria Baba, "Application of Independent Component Analysis to Handwritten Japanese Character Recognition", *Int. Joint Conf. on Neural Networks*, Washington, 1999.
[29] Kijsirikal Boonserm, Sinthupinyo Sukree and Supanwansa Apinya, "Thai Printed Character Recognition by Combining Inductive Logic Programming with Back Propagation Neural Network", *IEEE Proc. On Circuits and Systems - Micro Electronics and Integrating Systems*, Chiangmai, Thailand, 1998, pp. 539-542.
[30] Phokharaatkul Pisit and Kimpan Chom, "Recognition of Hand Printed Thai Characters Using the Cavity Features of Character Based on Neural Network", *IEEE Pro. On Circuit and Systems - Micro Electronics and Integrating Systems*, Chiangmai, Thailand, 1998, pp. 149-152. [31] Soo H. Kim, S. Jenog and Ching Y. Suen, "A Lexicon - driven Approach for Optimal Segment Combination in Off-line Recognition of Unconstrained Handwritten Korean Words", *Pattern Recognition*, Vol.34, 7, 2001, pp. 1137-1447.
[32] S.N.S. Raja Sekaran and B.L. Deekshatulu, "Recognition of Printed Telugu Characters", *Computer Graphics and Image Processing*, Vol. 6, 1977, pp. 335-366.
[33] M.B. Sukha Swamy, P. Seetharamulu and A.K. Pujari, "Recognition of Telugu Characters Using Neural Networks", *Int. J. of Neural Systems, Vol.* 6, 3, 1995, pp. 317-357.
[34] P.V.S. Rao and T.M. Ajitha, "Telugu Script Recognition - A Feature Based Approach", *Proc. of Third Int. Conf. on DAR*, Vol.1, 1995.
[35] Atul Nagi, Chakravarthy Bhagavathi and B. Krishna, "An OCR System for Telugu", *Proc. ICDAR*, Seattle, WA, 2001.
[36] Atul Nagi, K. Nikhil Shankar and Chandra Kanth Chereddi, "Localization, Extraction and Recognition of Text in telugu Document Images", *Proceedings of seventh International Conference on Document Analysis and Recognition(ICDAR 2003)*,2003.

[37] Arun K. Pujari, C.Dhanunjaya Naidu and B C Jinga, “An Adaptive Character Recogniser for Telugu Scripts Using Multiresolution Analysis and Associative Memory”,
[38] K. Sandhya Rani, Modular Neural Network Approach for Recognition of Printed Telugu characters, Ph.D thesis, 2003. SPMVV,Tirupati
[39] P. Sreevani, A Novel Approach for Recognition of Printed Telugu Document Using Neural Networks, Ph.D thesis, 2006,SPMVV, Tirupati
[40] Anbumani Subramanian and Bhadri Kubendran, “Optical Character Recognition of Printed Tamil Characters”, Department of Electrical and Computer engineering, Virginia Tech, Blacksburg. December 10, 2000.
[41] Ralph Niels and Louis Vuurpijl, “Dynamic TimeWarping Applied to Tamil Character Recognition**”** *Nijmegen Institute for Cognition and Information*
[42] H. C. Fernando, **“**A Two Stage Classification Approach to Tamil Handwriting Recognition” *Tamil Internet 2002, California, USA. 2002*
[43] G. Siromoney, R. Chandra Sekharan and M. Chandra Sekaran, "Machine Recognition of Printed Tamil Characters", *Pattern Recognition*, Vol. 10, 1978, pp. 243-247.
[44] M. Chandra Sekaran, R. Chandra Sekharan and G. Siromoney, "Context Dependent Recognition of Hand Printed Tamil Characters", *Proc. Int. Conf. System, Man and Cybernetics*, India, Vol. 2, 1984, pp. 786-790.
[45] R.M. Suresh, S. Arumugam, and L. Ganesan, "Recognition of Hand Printed Tamil Characters Using Fuzzy Finite Automata", *CSI 99 Conf*., Mumbai, 1999, pp. 1-4.
[46] R. Srinivas Rao Kunte and R.D. Sudhaker Samuel ,”On-Line Character/Script Recognition System for Kannada and Other Indian Languages Employing Wavelet Features”
[47] T.V. Ashwin and P.S. Sastry, " A Font and Size Independent OCR for Printed Kannada Documents Using SVM Classifier", *Proc. Of Open Research Forum, Fifth ICDAR*, Bangalore, India, 1999, pp. 14-18.
[48] A. Dutta and S. Chandhuri, "Bengali Alpha Numeric Character Recognition Using Curvature Features", *Pattern Recognition*, Vol. 26, 1993, pp. 1757-1770.
[49] A.K. Ray and B. Chatterjee, "Design of a Nearest Neighbour Classifier System for Bengali Character Recognition", *J. Inst. Elec. Telecom. Engg. (India)*, Vol. 30, 1984, pp. 226-229.
[50] B. Chaudhuri and U. Pal, "A Complete Printed Bangla OCR System", *Pattern Recognition*, Vol. 31, 1998, pp. 531-549.
[51] A.F.R. Rahman, "Recognition of Handwritten Bengali Characters: A Novel Multistage Approach", *Pattern Recognition*, Vol.35, 2002, pp. 997-1006.
[52] Sameer Antani and Lalitha Agnihotri, "Gujarati Character Recognition", *Fifth Int. Conf. Document Analysis and Recognition*, Bangalore (India), 1999, pp. 418.
[53] G.S. Lehal, Chandan Singh and Riha Lehal, "A Shape Based Post Processor for Gurumukhi OCR", *Int. Conf. Document Analysis and Recognition*, Seattle, WA, 2001.
[54] I.K. Sethi, "Machine Recognition of Constrained Hand Printed Devanagari", *Pattern Recognition*, Vol. 9, 1977, pp. 69-75.
[55] U. Pal and B.B. Chaudhuri, "Printed Devanagari Script OCR System", *Vivek*, Vol. 10, 1997, pp. 12-24.
[56] R.M.K. Sinha and H. Mahabala, "Machine Recognition of Devanagari Script", *IEEE Trans. System, Man and Cybernetics*, Vol. 9, 1979, pp.435-449.
[57] R.M.K. Sinha, "Rule Based Contextual Post Processing for Devanagari Text Recognition", *Pattern Recognition*, Vol. 20, 1987, pp. 475-485.
[58] Veena Bansali and R.M.K. Sinha, "On How to Describe Shapes of Devanagari Characters and Use Them for Recognition", *Proc. of Fifth Int. Conf. On Document Analysis and Recognition*, Bangalore, India, 1999, pp. 653-656.
[59] Veena Bansali, “Integrating Knowledge source in Devanagari Text Recognition”, *Ph.D thesis*, 1999.
[60] Huanfeng Ma and David Doermann, “**A**daptive Hindi OCR using generalized Hausdorff image comparison”, *ACM Transactions on Asian Language Information Processing (TALIP)* **,**Vol. 2(3), pp: 193 – 218, 2003 .
[61] B.Indira, M.Shalini, M V Ramana Muthy, Mahaboob Sharief Shaik “Classification and Rrecognition of Printed Hindi Characters Using ANNs” , in I.J. Image, Graphics & Signal Processing,2012, 6, 15-21.
[62] Aditi Goyal, Kartikay Khandelwal, Piyush Keshri “ Optical Character recognition for Handwritten Hindi” Stanford University.
[63] Divya Sharma “ Recognition of Handwritten Devnagari Script using Soft computing”, Thesis report, Master of Engineering, Thapar University.