

# Offline Recognition of Handwritten Urdu Characters using B Spline Curves: A Survey

Mohd Jameel  
School of Computer and  
System Sciences  
Jaipur National University, Jaipur

Sanjay Kumar  
School of Computer and  
System Sciences  
Jaipur National University, Jaipur

## ABSTRACT

Handwritten Character Recognition is an active area of research in the field of pattern recognition and image processing for last two decades as there is an urgent need of having a successful Script Recognition System to convert handwritten documents into computer understandable form which is applicable for various purposes. Several research studies have been carried out for recognition of other scripts like Chinese, Japanese, English, Devanagari, etc. but the research regarding Urdu Script is still immature due to cursive and variable nature of Urdu characters. The requirement of offline Urdu HCR systems is increasing because of the expansion of technology and the convenience for users. In this paper, a detailed survey of Urdu HCR techniques with respect to feature extraction developed so far alongwith their efficiency and accuracy has been presented. The paper also presents a new proposed B-Spline Curve approximation approach for feature extraction of offline isolated Urdu handwritten characters.

## Keywords

Handwritten, Urdu, Character, Recognition, B-Spline curve, Offline

## 1. INTRODUCTION

In the present era of digital revolution, it has become necessary to have all the available information in a digital form which is editable by the computers. In regions like Indian subcontinent, where there is a lot of information in the form of manuscripts, ancient books, etc. that are traditionally available in printed and handwritten form but are rarely available in digital form for searching. It has to be digitized and converted to the textual form in-order to be recognized by machines and searched quickly. In this regard, Script Recognition is an ultimate option to be adopted to achieve the desired goal. Character Recognition (CR) is one of the most important areas of pattern recognition and artificial intelligence. It is the process of detecting characters from input image and converting them into machine recognizable form. The main benefit of the Character Recognition process is that it can save both time and efforts when developing a digital replica of the document. The Character Recognition System is of two main types; Offline recognition and online recognition. In case of online Character recognition, the chronological sequence of points written through an input device such as stylus is traced out and pixel positions of characters is recorded instantly whereas in offline CR system, the written document image is scanned and processed.

Comparatively, online CR system development is usually easier than the offline as more information is available in the form of pixel coordinates. Offline character recognition system is to recognize what letters or words represent in a digital image of handwritten document. A lot of research has

been conducted in case of Printed Character Recognition whereas the Handwritten Character Recognition is still an open problem because of the different writing styles. Conversion of handwritten characters is important for making manuscripts into machine recognizable form so that it can be easily accessed and preserved. Many researchers have worked in the area of handwriting recognition and numerous techniques and models have been developed to recognize handwritten text.

## 2. CHARACTERISTICS OF URDU WRITING

Urdu is one of the constitutional languages of India and a very popular language of South Asia having a population of around 250 millions native speakers besides Middle East, USA and Europe. Urdu is an Indo-Aryan language of the Indo-Iranian branch and belongs to the Indo-European family. In the spoken form, Urdu and Hindi (together called Hindustani, the third main language of the world) are almost identical. However, in the written form, both languages are diametrically opposite. Urdu is in Arabic like script Whereas Hindi in the Devanagari script.

### 2.1 Urdu Characters and Numerals

The Urdu language consists of 38 basic characters and 10 numerals (Figure 1). Each character has an associated sound and most of them are multi-stroke characters and share a common structure. The main feature which distinguishes one character from another is the no. and position of “dots” and “toyeins”[1]. The positioning of the secondary strokes is also an important factor in distinguishing between different characters. The complexity further increases in case of handwritten Urdu characters as compared to printed ones.



Fig.1 Urdu Characters and Numerals

### 2.2 Peculiarity in Urdu Language Characters

The Urdu language is a cursive and context sensitive language. The characters are written from right to left but the digits / numbers are written from left to right. The characters consist of joiners and non-joiners that create 2-4 shapes for character depending upon its position/context in the

ligature/sub- word. The numbers/digits are free from the number of shapes regarding position.

- Some Urdu characters and numerals are misclassified and not recognized properly due to similar structures.
- Sometimes a digit confuses with other digit by model or system and is recognized wrongly.
- Some characters are not written properly as per language rules in handwritten case.

Keeping in view, the complexities of handwritten Urdu text and challenges involved in its recognition, the full fledged Handwritten Urdu Text Recognition System can be perceived as a combination of following milestones:

- Recognition of individual characters having single continuous structure.



Fig.2 One stroke characters

- Recognition of individual characters having two or three connected parts/components.

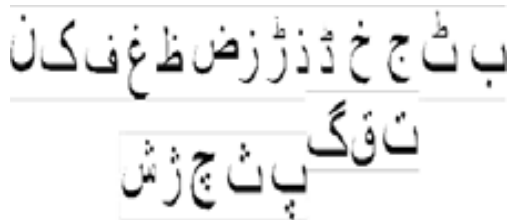


Fig.3 Two and three stroke characters

- Recognition of ligatures/words.
- Recognition of Sentences.

Each of the above components requires extensive research work to accomplish the desired task.

### 3. LITERATURE SURVEY

Handwritten Character Recognition system involves several major steps as outlined in Figure.4. These include preprocessing, segmentation, feature extraction, classification and post processing.

#### 3.1 Preprocessing

Preprocessing techniques have generally been found common in all handwritten script recognition systems. That is why a brief account of major preprocessing techniques found in literature is given here under:

**Thresholding:** This technique cited by [2,3] is used to transform an input image from gray scale to binary form to make it suitable for extracting most appropriate features.

**Thinning:** The thinning technique reduces thickness of the binary image of the character to one pixel as applied by Hanan et al [3] in their research.

**Normalization:** It is the process of resizing the image of the character to a standard sized matrix such as 30x30 or 40x40 pixels used by [4].

**Noise Removal:** It removes the unwanted bit patterns from the image which are not required in the output. Median filtering technique has been used by most of the researchers to remove the noise. Other techniques include Skeletonisation, skew detection, slant detection and correction and smoothening.

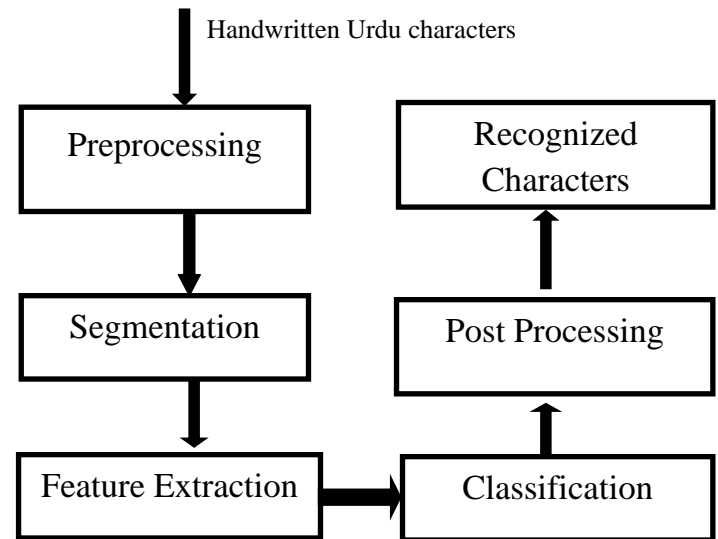


Figure 4. Stages of HCR

#### 3.2 Segmentation

Segmentation is the process of splitting the text into lines, words, characters or strokes. A typical Character Recognition System can have either holistic or character based approach (explicit or implicit) [5]. In holistic approach, the system recognizes the word or ligature directly without splitting it into characters. In segmentation or character-based approach, words or ligatures are segmented into characters or strokes implicitly or explicitly. Some segmentation techniques used for Urdu text are outlined in this paper. Javid et al [6] employed horizontal projection profile to segment the lines and separate ligatures from the diacritics while looking to the 8-neighbors. The association of dots and marks, with the relevant base forms, is carried out with the help of centroid-to-centroid distance and achieved 94% accuracy for a set of 1282 unique ligatures. The image signature scaling based method performing both horizontal and vertical scaling of Urdu, Arabic, and English text images for the signature calculation has been used by Azam et al. [7]. A segmentation based technique proposed in [8] for the recognition of Urdu script on compound ligatures handled the six classes of characters in single-column documents. The diacritics and main bodies are separated first and then thinning is applied. The ligatures are segmented only at outgoing directions. The freeman chain codes (FCC) technique has been applied for separation of primary and secondary ligatures and vertical scanning for primary ligatures extraction by [9].

The connected components are extracted in the binarized image of printed Urdu text by [10] to segment it into ligatures or partial words to which a set of two scalar and four vector features stored in the database represent.

#### 3.3 Feature Extraction

Feature extraction is the most crucial phase of any Character Recognition System and is a deciding factor for achieving a better recognition rate. A careful selection and extraction of features leads to enhancement in overall performance of the

recognition system. In this paper, a brief account of several feature extraction techniques used for handwritten Urdu character recognition in particular and other scripts in general has been given.

In earlier research works, Megherbi et al. [11] have used structural features. These include the number of dots present in the character, place of the dot, branch or presence of secondary stroke, aspect ratio and slope between the initial point and the end point. Structural features also play an active role in the technique employed in [12], in which visible features, such as the location of the dots and placement of other diacritics are extracted for every ligature. Structural features, such as character lengths, number and position of loops or holes, and distance between two consecutive lines have also been extracted by [13] in their research. The structural features are again important in the works of [14,15]. Another work in [16], the list of structural features includes loop, curve, cross, height of character, width of character, number of the dots, and position of the dots. In [17], the extracted features include height, width, and checksum from each character that differentiates one character from another. The width of a character is calculated by counting the black pixels and then calculating the difference between the first and the last black pixel from left to right. In the same way, the height is calculated from top to the bottom. The identification of Urdu ligatures based on the extraction of the statistical features, namely the axes ratio, solidity, eccentricity, moment based features, normalized length features, the number of holes, and the curvature features has been found in the work of [18]. The details regarding rotation, translation, and scaling invariant features are extracted from the base ligature into a set using the RTS invariant moment and then the extracted special ligatures are linked to the most probable neighboring base ligature using the centroid-to-centroid distance. The research conducted in [19] has extracted a combination of topological, contour, and water reservoir features for the individual characters of Urdu script. Lodhi et. al [20] have proposed a RTS invariant method that considers Fourier descriptors for the feature selection of Urdu characters. Fourier descriptors are used to uniquely represent the given characters' polygonal signatures. In [21], the global transformation method is applied for the extraction of features from a ligature prior to segmentation. Zaman et al [22] have used the row-major or the column-major order after applying various preprocessing methods for the conversion of a normalized image into a row vector of binary values as a feature input. Wavelet transformation features have been used by [23 and 24]. In [25 and 26] the researchers have tried to determine the body and secondary part, position of the part above or below, loop and Radon transform of the characters. Code of chain feature are used in [27] and Pseudo-Zernike Moments, size, rotation and translation invariant features have also been in [28].

Sagheer et al. [29] have made an attempt on the Urdu isolated handwritten digits in which they used Gradient features after developing the dataset on normalized images of digits. Basu et al. [30] presented recognition system for postal address code for Latin, Devanagari, Bangla and Urdu. Hough transformation is applied for localization and isolation and then digits from 4 scripts are clustered into 25 groups and extracted multiple features like Quad Tree Longest Run, overlapping longest-run, chain-code histogram and Gabor filter based features, shadow features, shadow-longest run-octant centroid and combinations of shadow- longest run. Razzak et al. [31] performed binarization, skew and slant correction and normalization in the offline domain. They used

combination of both structural features like holes, start and end point of the chain code, right, left, up and down direction of the pixel in a digits, number of strokes, number and position of cusp etc. for online Urdu and Arabic digit recognition. Some efforts on Arabic and Persian character recognition have been found but the work is not as matured as the Latin character recognition. There is still opportunity for the researchers to work for achieving the best character recognition system. Shokoohi et al [32] performed experiment of CNN on features extracted using non-linear algorithm from CENPARMI Persian dataset and achieved 78% recognition rate. Nooraliei [33] presented zoning and projection histogram features for handwritten Farsi numeral recognition. In 2014, Roy et al. [34] performed experiments on Arabic for handwritten numeral recognition and proposed Axiomatic Fuzzy Set theory for feature selection. Different kinds of features like Fourier switch features, stroke density features, contour features, projection features, the barycenter and barycenter distance feature were extracted from handwritten Arabic digits and then the features were filtered using under and outer analogy. Recently in 2014, Ghaleb et al. calculated density of digits through horizontal and vertical centroid moments and then used minimum distance classifier. Liu et al[35] have proposed a character stroke extraction method for handwriting recognition based on B-spline curve matching. They modeled the character as a set of B-splines, each of which represents a character stroke. They investigated the use of this method on data of handwritten English and handwritten Chinese characters and got effective results. In another attempt, Miura et al [36] have extracted curvature features based on a curve fitting approximation. They obtained cubic B-splines using a least squares method with natural conditions at the endpoints. The method was tested for English characters and numerals and acquired promising results. Zhongkang [37] proposed a rational B-spline representation of Handwritten digit templates based on Pixel-to-Boundary Distance maps for extracting and optimizing templates to develop a classifier that can reliably reject non digit patterns while achieving a high

Recognition rate on connected handwritten digit strings. Tirandaz et al [38] have presented a efficient technique for curve matching and character recognition. This technique is based on constructing and comparing B-spline Curves of object boundaries, calculation of the dominant points on boundary using Local Curvature Maximum and then control points are obtained by using B-spline least square fitting method. Character recognition is done by matching comparing resultant B-spline Curves. Some of the feature extraction techniques used in respect of handwritten character recognition systems alongwith accuracy are summarized in Table.1

#### **4. CLASSIFICATION**

Classification stage is also an important stage of any Character Recognition System in which the features extracted through feature extraction stage are fed as input to the model for identification and recognition. This paper mainly concentrates on feature extraction techniques used for the handwritten character recognition however; a brief overview of some classification models used for handwritten character recognition system for Urdu and some other scripts has been given. These include Hidden Markov Model (HMM), Support Vector Machines (SVM), Artificial Neural Networks (ANN), k-Nearest Neighbors (k-NN), fuzzy logic, genetic algorithm and others.

The recognition system in [16] has made use of the Kohonen self-organizing map(K-SOM) for the recognition of extracted and pre-segmented Urdu characters and 80% accuracy was achieved. . A hidden Markov model (HMM)is used for the classification of segmented primitives of the ligature by calculating the DCT features for improving the performance

of recognition [8]. The system mentioned in [18] trains a feed forward, back propagation, neural network model on a pre-defined set of ligatures from Urdu script. It gives good accuracy for 200 tested ligatures but not satisfactory results for unknown ligatures.

**Table.1 Comparison of handwritten Urdu isolated character, word and numeral recognition.**

Author	Type	Feature	Classification	Accuracy
Pathan et al. [39]	Individual Character	Moment Invariant	SVM	93.59%
Ali et al. [40]	Word	Curvature, slope & variance of stroke	Neural Networks	70%-80%
Mukhtar et al. [41]	Word	Gradient and structural	SVM	70%–82%
Sagheer et al. [42]	Word	Gradient and structural	SVM	97.00%
Yusuf and Haider [43]	Numeral	Shape context & Bending energy	Bipartite graph matching	unspecified
Sagheer et al. [44]	Numeral	Gradient	SVM	98.61%
Basu et al. [45]	Numeral	QTLR	SVM	96.2%
Razzak et al. [46]	Word	Fuzzy logic	HMM and Fuzzy Logic	89.2%
Husain et al. [47]	Word	Loop, intersection, writing styles of ligatures	Back Propagation Neural Network	93%

#### **4. PROPOSED TECHNIQUE BASED ON B-SPLINE CURVE FOR HANDWRITTEN URDU CHARACTER RECOGNITION**

After going through the literature, it has been found that the use of B-Spline Curve in recognition of Handwritten Urdu Characters has not yet been made (to our knowledge) whereas the application of B-Spline curves has shown better results in recognizing English and Chinese handwritten characters[35]. The benefit of using B-Spline Curves for recognition of Handwritten Urdu Characters lies in the fact that the B-Splines are the continuous curve representations and affine

invariant [50]. We know that the characters are formed by certain curves and hence each letter or character may be represented by a curve. Deing invariant under affine transformations such as translation, rotations, scales, B-Spline curve is suitable for simulating handwritten text.

We shall work on the proposed technique by taking individual Urdu characters, perform preprocessing, converting character image into binary and forming a bit matrix of the same, Skeletonise it, find out control or dominants points using an appropriate method such as Slalom method and then constructing the B-Spline curve to investigate the results using the samples of Handwritten Characters of Urdu script from different writers and also from CENPARMI-U database.

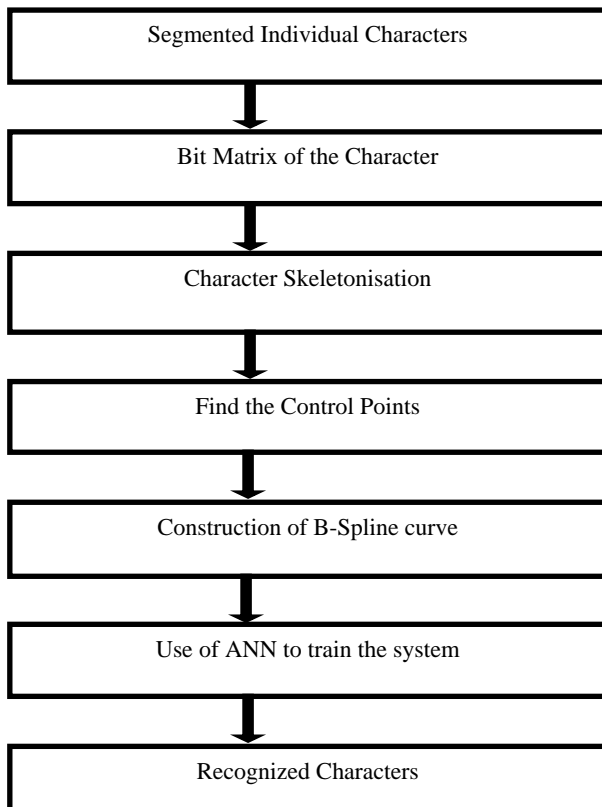


Figure.5 Main steps of Urdu HCR using B-Spline Curves

## 5. CONCLUSION

The aim of this paper was to provide a detailed survey of published research work in the sequential stages of Urdu handwriting character recognition with special attention towards feature extraction techniques. We have found that various characteristics such as feature extraction, preprocessing, segmentation and recognition techniques have been used and reported different accuracy levels but the use of B-Spline Curve has not been found for handwritten Urdu script to form the feature vector in spite of their robustness. We have proposed a technique in this regard in order to enhance the accuracy and efficiency of Urdu HCR.

## 6. REFERENCES

- [1] Shahzad N, Paulson B and Hammond Tracy "Urdu Qaeda: Recognition System for Isolated Urdu Characters" Sketch Recognition Lab. Texas A&M University, 2009.
- [2] H. Aljuaid and D. Muhamad, "Offline Arabic Character Recognition using Genetic Approach: A Survey".
- [3] R. I. Zaghoul, E. F. Alrawashdeh, D. Mohammad, and K. Bader, "Multilevel Classifier in Recognition of Handwritten Arabic Characters" Journal of Computer Science 7 (4): 512-518, ISSN 1549-3636, 2011.
- [4] H. Aljuaid, Z. Muhammad and M. Sarfraz, "A Tool to Develop Arabic Handwriting Recognition System Using Genetic Approach", Journal of Computer Science 6 (6): 597- 602 ISSN 1549-3636, 2010.
- [5] Saeeda Naz, Arif Iqbal, Umar "An OCR System For Printed Nasta'liq Script: A Segmentation Based Approach" ISBN: 978-1-4799-5754-5/14/©2014 IEEE.
- [6] S.T.Javed, S.Hussain "Improving Nastalique-specific pre-recognition process for Urdu OCR" Proceedings of the 13th International Multi-topic IEEE Conference (INMIC'09), 2009, pp.1-6.
- [7] S.M.Azam, Z.A.Mansoor, M.Sharif, "On fast recognition of isolated characters by constructing character signature database" Proceedings of the International Conference on Emerging Technologies (ICET'06), 2006, pp.568-575.
- [8] S.T. Javed "Investigation into a segmentation based OCR for the Nastaleeq writing system (Master's thesis)". National University of Computer & Emerging Sciences, Lahore, Pakistan, 2007.
- [9] H. Malik, M.A. Fahiem, "Segmentation of printed Urdu scripts using structural features" Proceedings of the 2nd International Conference in Visualisation (VIZ'09), 2009, pp.191-195.
- [10] A.Abidi, I.Siddiqi, K.Khurshid "Towards searchable digital Urdu libraries-a word spotting based retrieval approach" Proceedings of the International Conference on Document Analysis and Recognition (ICDAR'11), 2011, pp. 1344-1348.
- [11] D.B. Megherbi, S.M.Lodhi, A.J.Boulenouar "Fuzzy-logic-model-based technique with application to Urdu character recognition" Proceedings of the SPIE Applications of Artificial NN in Image Processing V3962 (2000) 13-24.
- [12] Z.A.Shah "Ligature based optical character recognition of Urdu-Nastaleeq font" Proceedings of the 6th International Multi-topic IEEE Conference (INMIC'02), 2002, pp.25-25
- [13] Z.Ahmad, J.K.Orakzai, I.Shamsher, A.Adnan "Urdu Nastaleeq optical character recognition" Proceedings of the World Academy of Science, Engineering and Technology, vol.26, 2007.
- [14] S.A.Sattar, S.Haque, M.K.Pathan, Q.Gee "Implementation challenges for nastaliq character recognition" Wireless Networks, Information Processing and Systems, Communications in Computer and Information Science, vol.20. Springer, Berlin, Heidelberg, 2009, pp.279-285.
- [15] S.A.Sattar, S.ulHaq, M.K.Pathan "A finite state model for Urdu nastalique optical character recognition" International Journal of Computer Science and Network Security (IJCSNS) 9(9) (2009).
- [16] S.A. Hussain, S.Zaman, M.Ayub "A self organizing map based Urdu Nasakh character recognition" Proceedings of the International Conference on Emerging Technologies (ICET'09), Islamabad, Pakistan, 2009, pp.267-273.
- [17] J. Tariq, U.Nauman, M.U.Naru "Soft converter: a novel approach to construct OCR for printed urdu isolated characters" Proceedings of the 2nd International Conference on Computer Engineering and Technology (ICCET'10), vol.3, Singapore, 2010, pp.V3-495-V3-498.
- [18] S.A.Husain "A multi-tier holistic approach for Urdu Nastaliq recognition" Proceedings of the 6th International Multi-topic IEEE Conference (INMIC'02), 2002, pp.528-532.
- [19] U.Pal, A.Sarkar "Recognition of printed Urdu script" Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR2003), 2003, pp.1183-1187.

- [20] S.M. Lodhi, M.A.Matin "Urdu character recognition using Fourier descriptors for optical networks" Proceedings of the Photonic Devices and Algorithms for Computing VII, vol. SPIE 5907, 2005.
- [21] S.T. Javed, S. Hussain, A. Maqbool, S. Asloob, S. Jamil, H. Moin "Segmentation Free Nastalique Urdu OCR" World Academy of Science, Engineering and Technology
- [22] S.Zaman, W.Slany, F.Sahito "Recognition of segmented Arabic/Urdu characters using pixel values as their features" Proceedings of the 1st International Conference on Computer and Information Technology (ICCIT'2012), 2012.
- [23] N. B. Amor and N. E. Ben Amara, "Combining a hybrid Approach for Features Selection and Hidden Markov Models in Multi-font Arabic Characters Recognition", IEEE Second International Conference on Document Image Analysis for Libraries (DIAL'06) 0-7695-2531-8/06, 2006.
- [24] N. Ben Amor, M. Zarai, and N. E. Ben Amara, "Neuro-Fuzzy approach in the recognition of Arabic Characters" 0-7803-9521-2/06/\$20.00 ©2006 IEEE.
- [25] R. I. Zaghoul, E. F. Alrawashdeh, D. Mohammad, and K. Bader "Multilevel Classifier in Recognition of Handwritten Arabic Characters" Journal of Computer Science 7 (4): 512-518, ISSN 1549-3636, 2011.
- [26] G. A. Abandah, K. S. Younis and M. Z. Khedher, "Handwritten Arabic Character Recognition Using Multiple Classifiers Based on Letter Form" Proc. 5th IASTED Int'l Conf. on Signal Processing, Pattern Recognition, & Applications (SPPRA, Innsbruck, Austria, 2008.
- [27] O. Hachour "The Combination of Fuzzy Logic and Expert System for Arabic Character Recognition" 3rd International IEEE Conference Intelligent Systems, September 2006.
- [28] M. N., K. Faez "Recognition of Multi-font Farsi / Arabic Characters Using a Fuzzy Neural Network" IEEE.
- [29] M. W. Sagheer, C. L. He, N. Nobile, and C. Y. Suen "A New Large Urdu Database for Off-Line Handwriting Recognition" in Proceedings of International Conference on Image Analysis and Processing (ICIAP'09), 2009.
- [30] S. Basu, N. Das, R. Sarkar, M. Kundu, M. Nasipuri, and D. K. Basu "A Novel Framework for Automatic Sorting of Postal Documents with Multi-Script Address Blocks" Pattern Recognition., vol. 43, no. 10, pp. 3507–3521, 2010.
- [31] M. I. Razzak, S. A. Hussain, A. Belaïd, M. Sher, and others, "Multi-font Numerals Recognition for Urdu Script based Languages" Int. J. Recent Trends Eng., 2009.
- [32] Z. Shokoohi, A. M. Hormat, F. Mahmoudi, and H. Badalabadi "Persian handwritten numeral recognition using Complex Neural Network and non-linear feature extraction" First Iranian Conference on Pattern Recognition and Image
- [33] A. Nooraliei "Persian handwritten digits recognition by using zoning and histogram projection" AI & Robotics and 5th Robo Cup Iran Open International Symposium (RIOS) 3rd Joint Conference of, 2013, pp. 1–5.
- [34] A. Roy, N. Das, R. Sarkar, S. Basu, M. Kundu, and M. Nasipuri "An axiomatic fuzzy set theory based feature selection methodology for handwritten numeral recognition" Proceedings of the 48th Annual Convention of Computer Society of India-Vol I, 2014, pp. 133–140.
- [35] Xiabi Liu and Yunde Jia "Character Stroke Extraction Based on B-spline Curve Matching by Constrained Alternating Optimization" Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)
- [36] K. T. Miura, R. Sato and S. Mori "A Method of Extracting Curvature Features and its Application to Handwritten Character Recognition" 0-8186-7898-4/97 /1997 IEEE.
- [37] Zhongkang "Extraction and Optimization of B-Spline PBD Templates for Recognition of Connected Handwritten Digit Strings" IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 1, January 2002.
- [38] H. Tirandaz, A. Nasrabadi and J. Haddadnia "Curve Matching and Character Recognition by Using B-Spline Curves" IACSIT International Journal of Engineering and Technology Vol.3, No.2, April 2011
- [39] I.K. Pathan, A.A.Ali "Recognition of offline handwritten isolated Urdu character" Advances in Computational Research 4(1)(2012)117–121.
- [40] A.Ali, M. Ahmad, N.Rafiq, J.Akber, U.Ahmad, Akmal "Language independent optical character recognition for handwritten text" Proceedings of the 8th International Multi-topic IEEE Conference (INMIC'04), 2004, pp.79–84.
- [41] O. Mukhtar, S.Setlur, V.Govindaraju "Experiments on Urdu text recognition" Guide to OCR for Indic Scripts, Advances in Pattern Recognition. Springer, London, 2010, pp.163–171
- [42] M.W.Sagheer, C.L.He, N.Nobile, C.Y.Suen "A new large Urdu database for off-Line handwriting recognition"
- [43] M. Yusuf, T.Haider "Recognition of handwritten Urdu digits using shape context" Proceedings of the 8th International Multi-topic IEEE Conference (INMIC'04), 2004, pp.569–572.
- [44] M.W.Sagheer, C.L.He, N.Nobile, C.Y.Suen "Holistic Urdu handwritten word recognition using support vector machine" Proceedings of the 20th International Conference on Pattern Recognition (ICPR'10), 2010, pp.1900–1903.
- [45] S. Basu, N.Das, R.Sarkar, M.Kundu, M.Nasipuri, D.K.Basu "A novel framework for automatic sorting of postal documents with multi-script address blocks" Pattern Recognition 43(10)(2010)3507–3521.
- [46] M.I. Razzak, F.Anwar, S.A.Husain, A.Belaïd, M.Sher "HMM and fuzzy logic: a hybrid approach for online Urdu script-based languages' character recognition" Knowledge Based Systems 23(8)(2010)914–923.

- [47] S.A. Husain, A.Sajjad, F.Anwar “Online Urdu character recognition system” Proceedings of the IAPR Conference on Machine Vision Applications (MVA'07), 2007,pp.98–101.
- [48] Saeeda Nazl, Saad B. “Arabic Script based Digit Recognition Systems” International Conference on Recent Advances in Computer Systems (RACS 2015)
- [49] S. Naz, et al. “The optical character recognition of Urdu-like cursive scripts” Pattern Recognition (2013), <http://dx.doi.org/10.1016/j.patcog.2013.09.037i>
- [50] Khoi Nguyen-Tan and Nguyen Nguyen-Hoang “Handwriting Recognition Using B-Spline Curve” ICCASA 2012, LNICST 109, pp. 335–346, 2013.