CrossMark

Character and numeral recognition for non-Indic and Indic scripts: a survey

Munish Kumar[1] · **M. K. Jindal**[2] · **R. K. Sharma**[3] ·
Simpel Rani Jindal[4]

Abstract A collection of different scripts is employed in writing languages throughout the world. Character and numeral recognition of a particular script is a key area in the field of pattern recognition. In this paper, we have presented a comprehensive survey on character and numeral recognition of non-Indic and Indic scripts. Many researchers have done work on character and numeral recognition from the most recent couple of years. In perspective of this, few strategies for character/numeral have been developed so far. There are an immense number of frameworks available for printed and handwritten character recognition for non-Indic scripts. But, only a limited number of systems are offered for character/numeral recognition of Indic scripts. However, few endeavors have been made on the recognition of Bangla, Devanagari, Gurmukhi, Kannada, Oriya and Tamil scripts. In this paper, we have additionally examined major challenges/issues for character/numeral recognition. The efforts in two directions (non-Indic and Indic scripts) are reflected in this paper. When compared with non-Indic scripts, the research on character recognition of Indic scripts has not achieved that perfection yet. The techniques used for recognition of non-Indic scripts may be used for

✉ Munish Kumar
  munishcse@gmail.com

  M. K. Jindal
  manishphd@rediffmail.com

  R. K. Sharma
  rksharma@thapar.edu

  Simpel Rani Jindal
  simpel_jindal@rediffmail.com

1 Department of Computer Applications, GZS Campus College of Engineering and Technology, Bathinda, Punjab, India

2 Department of Computer Science and Applications, Panjab University Regional Centre, Muktsar, Punjab, India

3 Department of Computer Science and Engineering, Thapar University, Patiala, Punjab, India

4 Computer Science and Engineering, Yadavindra College of Engineering, Talwandi Sabo, Bathinda, Punjab, India

🙌 Springer

recognition of Indic scripts (printed/handwritten text) and vice versa to improve the recognition rates. It is also noticed that the research in this field is quietly thin and still more research is to be done, particularly in the case of handwritten Indic scripts documents.

Keywords OCR · Character recognition · Non-Indic scripts · Indic scripts

1 Introduction

Character and numeral recognition frameworks have been a point of research for as far back as a couple of decades. Still, it remains an exceptionally difficult task to implement a character and numeral recognition framework that works under each possible condition and gives very precise outcomes. In Optical Character Recognition (OCR), the patterns are alphabets and numbers and so forth, while the diverse classes corresponding to the distinctive characters. The instructing of the machine is performed by demonstrating the machine cases of characters of all the diverse classes. In light of these cases the machine constructs a model of each class of characters. At that point, unknown pattern (character or number) is compared with the beforehand acquired depictions, and relegated the class that gives the best match. Optical recognition is performed after the writing of text or printing on paper has been finished, rather than on-line recognition where the computer system recognition recognizes the characters as they are drawn. Both handwritten and printed characters may be recognized accurately, but the quality is straightforwardly needy upon the superiority of the documents. The research in the area of character recognition began in the nineteenth century and first optical character recognition was on hand in 1929. Modern version of OCR was produced in 1951 by David Shepard (Schantz 1982). Character recognition has played and currently playing an important role in pattern recognition research. In general, research on optical character recognition for Indic scripts is in progress. But till now, no solution has been offered that solves the problem correctly and efficiently for Indic scripts. There are various applications of character/numeral recognition research like handwritten notes reading, banking cheque reading, post code recognition, form processing, etc. Character recognition system can be used for reading handwritten notes. Notes are, normally, used to record facts, topics, or thoughts, written down as an assist to memory. Cheque reading is a very important commercial application of character recognition system. Character recognition system plays a very important role in banks for signature verification and for recognition of amount filled by the user. Character and numeral recognition system can be used for reading handwritten postal address on letters and handwritten digits of postcodes. Character recognition system can also be used for form processing. Forms are normally used to collect information from the public. This information can be processed by using a handwritten character recognition system. Signature identification is the specific field of handwriting OCR in which the writer is verified by some specific handwritten text. Offline handwritten character recognition system can be used to identify a person by handwriting, as handwriting varies from person to person. This paper consists of five sections. In Sect. 2, we have presented various issues and challenges for character recognition. Section 3 presents the various motivations related to this work. In Sect. 4, the recognition of different non-Indic scripts has been reviewed and in Sect. 5, literature review has been conducted for Indic scripts. Section 6 presents recognition accuracies achieved for typical non-Indic and Indic scripts and in Sect. 7, we have discussed about a few suggestions on future directions of character recognition of different scripts. Finally, in Sect. 8, we have concluded this paper.

2 Challenges and issues

Various challenges are identified which may provide more lively interest to the researchers for character/numeral recognition. These challenges are difficult to identify the various styles of human writing, different shapes and size of letter, pure input quality, low accuracy rate in recognition etc. Hence, a lot of research work is to be done to solve these problems. Quality of the input document and varying font styles is a challenging task for various scripts. As compared to non-Indic scripts, Indic scripts have many additional challenges like larger character set due to modifiers, lack of standard test databases etc. Segmentation is a major challenge in text recognition due to existence of different skew angles between lines on the page or even along the same text line, i.e., presence of multiple skew in a document complicates the line segmentation process. Curvilinear lines and fluctuating lines also create problems while identifying the exact line boundaries, difference in skew angle of words and characters within the same line also hinders the process of line segmentation, overlapping adjacent text lines, i.e., lower or upper portion of one line extended to neighboring text line makes line segmentation difficult. Poor quality document consisting of holes, noise, spots, broken strokes etc. makes the process of line segmentation extremely difficult. Character segmentation poses additional problems due to touching, broken and overlapping characters. Non-uniform background is one of the quite challenging tasks for text recognition. Recognition of historical manuscript documents is also a challenging problem due to low quality of manuscripts, absence of standard alphabets, presence of unknown fonts, etc. Various challenges are identified for Arabic text recognition as Arabic is cursive in nature because individual characters join together to form a complete word, thus identifying the segmentation points in the words to separate isolated character becomes difficult. One of the major challenges for Arab is that Arabic characters are enriched with dots and diacritics, relative position of dots and diacritics changes frequently with respect to the character which it is associated. Sometimes, Japanese text recognition is also a challenging task to determine whether two radicals are in fact two separate characters or two component parts of the same character. Still, the largest challenge is recognizing the large number of characters and the majority of research has been devoted to overcoming this difficulty. Few major challenges and issues for Japanese text recognition are cursive characters, word spotting, and a document image may consist of printed and handwritten text. Bangla character recognition is a great challenge for researchers because of the large number of characters, change of shape in word and in conjunctive characters. Recognition of the printed Devanagari script is the challenging problem since there is a difference in the same character due to diverse font family, font size, font orientation etc. Sometime same font and size may also have bold face character as well as normal ones. Thus, the width of the stroke is also an issue that interrupts recognition. There are few major challenges in Gurmukhi text recognition. In online Gurmukhi handwriting recognition challenges like confusing strokes, reverse handwriting, new classes in handwritten words etc. are exiting. In offline handwriting recognition, sometimes headlines of the words are not straight, many touching or overlapping characters may be found in a word, and shape variation in different occurrences of a single character are the major challenges for Gurmukhi text recognition. There are few major challenges for Kannada text recognition are existing like, Kannada character set is very vast, few characters are similar with each other, size of characters and words in Kannada is not uniform.

3 Motivation

The advancement in optical character recognition framework is divided into two categories according to the technique for data acquisition: online character recognition and offline char-

acter recognition. The online character recognition framework utilizes the digitizer which particularly captures writing with the order of the strokes, speed, pen up and pen down data. Offline character recognition captures the information from paper through an optical scanner or cameras. Offline character recognition is otherwise called optical character recognition in light of the fact that the image of text is changed over into a bit design by optically digitizing devices. Recognition is carried out on this bit design information for both printed and handwritten text. Offline handwritten character recognition is difficult rather than online handwritten character recognition because stroke information is not available in offline handwriting. The major difficulties, as on account of any handwritten character recognition issue, are the huge variety in the composition styles of an individual at various circumstances and among various people, for example, shape, speed of composing and thickness of characters and so on. The issue of printed character recognition is generally solved and comprehended with few limitations and accessible framework yield approximately 99% recognition accuracy. But, handwritten character recognition has still constrained capacities. Other challenges incorporate the similarities of a few characters with each other, vast assortment of character shape and so forth. Offline handwritten character recognition is a standout amongst the most famous zones of research in document analysis and recognition of its enormous application potential. Offline handwritten character recognition is generally developed in scripts like Arabic, Chinese, Korea, and Roman. Some encouraging research findings are recorded in Indic scripts like Bangla, Devanagari and so on. In Indic scripts, though many research papers are published around there, the outcomes detailed are deficient for the outline of efficient handwritten character recognition frameworks. This is the inspiration driving this paper.

4 Recognition of non-Indic scripts

Borovikov (2004) have presented a survey of modern optical character recognition techniques. In this article, they discussed about the latest advances and major developments for optical character recognition techniques. Hussain et al. (2015) have presented a comprehensive survey of handwritten document benchmarks. They have also presented a comparison of these databases on a number of dimensions. The ground truth information of the database along with the supported tasks is also discussed by them. Sonkusare and Sahu (2016) have presented a survey on handwritten character recognition techniques for English alphabets. They have presented an outline of current research work conducted for recognition of handwritten English alphabets. A variety of recognition methodologies with their performance for handwritten English alphabets are conferred by them in this paper. Modi and Parikh (2017) have presented a detailed review in the field of optical character recognition. They have surveyed various techniques for pre-processing and segmentation phases on optical character recognition. We have noticed that most of the existing efforts on optical character recognition deal with non-Indic scripts. In the following sub-sections, we have presented literature of different non-Indic scripts.

4.1 Arabic

The Arabic script is used for writing Arabian and Persian languages. Almuallim and Yamaguchi (1987) have presented a recognition system for Arabic script. They have used geometrical and topological features for recognition. Impedovo and Dimauro (1990) have proposed a recognition system for handwritten Arabic numerals based on Fourier descriptors. Roy et al. (2004) have presented an Arabic postal automation system for sorting of

postal documents. Multi-Layer Perceptron (MLP) classifier has been reflected in their work for recognition of Bangla and Arabic numerals. They have obtained maximum recognition accuracy of about 92.1% for handwritten numerals. Lorigo and Govindaraju (2006) have presented a critical review on offline Arabic handwriting recognition systems. They have presented various techniques employed at different stages of the offline handwritten Arabic character recognition system. Izadi et al. (2006) addressed the issues in the Arabic alphabet, adopted and evolved, for writing Persian language. Abd and Paschos (2007) have obtained a recognition accuracy of 99.0% with Support Vector Machine (SVM) for the Arabic script. Alaei et al. (2009) have presented fivefold cross validation technique based recognition system for Arabic numerals. They have achieved a recognition accuracy of 99.4% on a 10-class problem with 20,000 samples in testing data set. Alaei et al. (2010a) have proposed a technique for segmentation of handwritten Persian script text lines into characters. The proposed algorithm finds the baseline of the text image and straightens it. They have extracted features using histogram analysis and removed segmentation points, using baseline dependent as well as language dependent rules. They have achieved maximum segmentation accuracy of 92.5%. Alaei et al. (2010b) have proposed a Persian isolated handwritten character recognition system. They employed SVM for classification and achieved a recognition accuracy of 98.1% with modified chain code features. Kacem et al. (2012) have used structural features for recognition of Arabic names. Shahin (2017) has introduced a system for printed Arabic text recognition using linear and nonlinear regression. He has tested his proposed methodology with 14,000 different words of Arabic script and accomplished a recognition accuracy of 86.0%. Althobaiti and Lu (2017) have presented a review on Arabic optical character recognition and they have proposed a technique for isolated handwritten Arabic character recognition based on encoded freeman chain code.

4.2 Chinese

Some Chinese character recognition systems based on the orthogonal moment descriptors have been reported (Liu and Ma 1996; Zhang et al. 1990; Yap and Paramesran 2003). Zhu et al. have recognized Chinese characters based on stroke and structural features. Liao wt al. (2002) have presented a method based on Gegenbauer moment for Chinese character recognition. Their method can provide a modest improvement in terms of recognition for those Chinese characters that are very similar in shapes. They have used a set of 6763 Chinese characters is used as the testing images. Das and Banerjee (2015) have presented an algorithm based on geometry topology for Japanese Hiragana character recognition. They have achieved an average recognition rate of 94.1%. Bluche and Messina (2016) have presented a technique for recognition of segmentation-free methods for handwritten Chinese text. He and Hu (2016) have presented a system for Chinese character recognition from natural scenes. They have presented a novel method based on the integrated channel feature and pooling technology to extract informative features from scenes images.

4.3 French

Tran et al. (2010) have considered the problem of French handwriting recognition using 24,800 samples. They have worked on both online and offline handwritten character recognition. Grosicki and Abed (2009) proposed a French handwriting recognition system in a competition held in ICDAR-2011. In this competition, they have presented comparisons between different classification and recognition systems for French handwriting recogni-

tion. Swaileh et al. (2016) introduced a new unified syllabic model for French handwriting recognition based on hidden Markov models (HMM).

4.4 Japanese

Nakagawa et al. (2005) have presented a model for online handwritten Japanese text recognition which is free from line direction constrains and writing format constraints. Zhu et al. (2010) have described a robust model for online handwritten Japanese text recognition. They received a recognition accuracy of 92.8% using 35,686 samples. Tsai (2016) has achieved a recognition accuracy of 96.1% for Handwritten Japanese Characters which consist of three different types of scripts: hiragana, katakana, and kanji. They have used deep Convolutional Neural Networks for classification. For experimentation work, they have considered Electrotechnical Laboratory (ETL) Character Database from the National Institute of Advanced Industrial Science and Technology (AIST). Liang et al. (2016) have presented an on-line handwritten Japanese text recognition system. They introduced a new unified syllabic model for French and English handwriting recognition, based on hidden Markov models (HMM). Their proposed method sets each off-stroke between real strokes as undecided and evaluates the segmentation probability by SVM model.

4.5 Roman

Schomaker and Segers (1999) have proposed a technique for cursive Roman handwriting recognition using geometrical features. Park et al. (2000) have presented a hierarchical character recognition system for achieving high speed and accuracy by using a multi-resolution and hierarchical feature space. They obtained a recognition rate of about 96.0%. Wang et al. (2000) have presented a technique for recognition of Roman alphabets and numeric characters. They had a recognition rate of about 86.0%. Bunke and Varga (2007) have reviewed the state of the art in offline Roman cursive handwriting recognition. They identified the challenges in Roman cursive handwriting recognition. Liwicki and Bunke (2007) have combined online and offline Roman handwriting recognition systems using a new multiple classifier system. They obtained a maximum recognition accuracy of 66.8% for the combination of online and offline handwriting recognition. Schomaker (2007) has presented a method for retrieval of handwritten lines of text in historical administrative documents. Chanda et al. (2007a) have proposed a SVM based method for identification of printed Roman script documents. They have extracted structural features for script identification and achieved 99.4% recognition accuracy. Pal et al. (2010) have proposed a bi-lingual city name recognition system for Bangla and English. They have considered 11,875 samples for testing and obtained 92.2% recognition accuracy. Jayadevan et al. (2010) have evolved a scheme for recognition of words used to write the amount of bank cheques. They collected a database of 5400 words from fifty writers for testing. Recognition accuracy of 97.0% has been achieved by them. Afroge et al. (2016) have proposed an optical character recognition system for Roman script using a back propagation neural network. They trained their network with more than 10 samples per class and give accuracy of 99.0, 97.0, 96.0 and 93.0% for numeric digits, capital letters, small letters and alphanumeric characters, respectively.

4.6 Thai

Chanda et al. (2007b) have evolved a method based on SVM for identification of printed Thai script documents. They have obtained 99.4% script identification accuracy. Karnchanapusakij

et al. (2009) have used linear interpolation approach for online handwritten Thai character recognition. They have obtained 90.9% recognition accuracy using this system. Kobchaisawat and Chalidabhongse (2015) have proposed a method for multi-oriented Thai text localization in natural scene images. They have considered convolutional neural network for classification. Asavareongchai and Giarta (2016) have presented an image processing system for recognition of Thai characters and text from documents. Sopon et al. (2017) have proposed a framework for Thai text retrieval using speech. They have achieved average word accuracy of 74.50%.

5 Recognition of Indic scripts

Pal et al. (2012) have presented a state-of-the-art survey about the techniques available in the area of offline handwriting recognition (OHR) in Indian regional scripts. They have presented survey of nine regional scripts and then categorized these nine scripts into four subgroups based on their similarity and evolutionary information. Various feature extraction and classification techniques associated with the offline handwriting recognition of the regional scripts are discussed in this survey. They have also discussed about the details of the datasets available in different Indian regional scripts. Singh et al. (2012) have presented a survey based on various applications of optical character recognition in different fields. Prasad (2014) has presented an in-depth literature survey of Indic script recognition systems for Bangla, Devnagari, Gurumukhi, Kannada, Malayalam, Tamil, and Urdu. They focused on a multitude of feature and classification techniques for recognition of various scripts. Koundal et al. (2017) have presented a survey for Punjabi character recognition. They have discussed about various feature extraction techniques and classification techniques explored for printed and handwritten Punjab character recognition. As compared to non-Indic scripts, the research on character recognition of Indic scripts has not achieved that perfection yet. So, research in the field of character recognition of Indic scripts is ongoing. In Indic scripts, there is mainly resolute character recognition of machine printed text. Limited attempts have been made for recognition of degraded printed text and handwritten text as well.

5.1 Bangla

A good number of researchers have worked for recognition of handwritten characters in Bangla script. Bangla script is used for writing Bengali and Assamese languages. Dutta and Chaudhury (1993) have presented a system for isolated Bangla alphabets and numerals recognition using curvature features. Pal and Chaudhuri (1994) have proposed a character recognition system using tree classifier. Their system was quite fast because pre-processing like thinning is not required in their scheme. They have achieved a recognition accuracy of 96.0% using 5000 characters data set. Bishnu and Chaudhuri (1999) have used a recursive shape based technique for segmentation of handwritten Bangla script documents. Pal and Dutta (2003) have proposed a system for segmentation of unconstrained Bangla handwritten connected numerals. They achieved segmentation accuracy of 94.8%. Roy et al. (2004) have presented a handwritten numeral recognition system for Indian postal automation and achieved a recognition accuracy of 92.1%. Bhattacharya et al. (2006) have presented Bangla character recognition system and they have obtained maximum recognition accuracy of 94.7%. Pal et al. (2006) have proposed a technique for slant correction of Bangla characters based on Modified Quadratic Discriminant Function (MQDF). They have achieved a recognition accuracy of 87.2% for Bangla city name images dataset. Bhattacharya et al. (2007) have presented an approach for online Bangla handwritten character recognition. They

developed a 50-class recognition problem and achieved an accuracy of 92.9 and 82.6% for training and testing, respectively. Pal et al. (2007a) dealt with recognition of offline handwritten Bangla compound characters using MQDF. They have obtained 85.9% recognition accuracy by using fivefold cross validation technique. Pal et al. (2008) have proposed a technique for Bangla handwritten pin code recognition system. Reddy et al. (2012a, b) have presented a handwritten numeral recognition system that can be employed to both online and offline situations for Assamese language. For online handwritten numeral recognition, they have used x and y coordinates for feature extraction and HMM classifier for recognition. For offline numeral recognition, they have considered projection profile features, zonal discrete cosine transforms, chain code histograms and pixel level features and Vector Quantization (VQ) classifier for recognition. They have achieved a recognition accuracy of 96.6 and 97.6% for online and offline handwritten numerals, respectively. Reddy et al. (2012a, b) have also presented a HMM based online handwritten digit recognition system using first and second order derivatives at each point as features. They obtained a recognition accuracy of 97.1% on 18,000 samples testing data set. Sarma et al. (2013) have presented handwritten Assamese numeral recognition system using HMM and SVM classifiers. They have attained a recognition accuracy of 96.5 and 96.8% with HMM and SVM classifier, respectively. Afroge et al. (2016) have presented an offline printed optical character recognition system based on multilayer perceptron model for Bangla script. They have proposed a feature extraction technique based on "Discrete Frechet Distance" and "Dynamic Time wrapping". They have achieved a recognition accuracy of 95% all basic characters of Bangla script.

5.2 Devanagari

The Devanagari script is used for writing four languages, namely, Hindi, Marathi, Nepali and Sanskrit. Sethi and Chatterjee (1976) have done a good amount of work on Devanagari script recognition. They have used binary decision tree classifier for recognition. Pal and Chaudhuri (2001) have proposed a methodology for machine recognition of printed and handwritten texts of Devanagari script. Recognition accuracy of 98.3% has been achieved by them using this system. Bansal and Sinha (2000) have also presented two phases based recognition system for Devanagari script. In the first phase, they recognized the unknown stroke and in the second phase, they identified the character based on strokes recognized in the first step. Roy et al. (2004) have evolved a scheme for handwritten script identification system and they have generated a tree classifier for word by word script identification of Bangla, Devanagari, and English. They have achieved a recognition accuracy of 98.4% with the proposed technique for printed text. Joshi et al. (2005) have presented an online handwritten Devanagari character recognition system. They have proposed structural feature based algorithm for recognition. Hanmandlu et al. (2007) have used membership functions of fuzzy sets for handwritten Devanagari script recognition. Pal et al. (2007b) have developed a modified classifier based scheme for offline handwritten numerals recognition of six widely used Indian scripts. They have extracted directional features for numeral recognition. They have obtained 99.6% recognition accuracy with fivefold cross validation technique. Pal et al. (2007d) have set into motion, a system for offline handwritten Devanagari character recognition. They have achieved a recognition accuracy of 94.2% with fivefold cross validation test. Kumar (2008) has brought in an artificial intelligence based technique for machine recognition of handwritten Devanagari script. He has used three levels of abstraction to describe this technique. Pal et al. (2009) have assimilated a comparative study of handwritten Devanagari character recognition. Garg et al. (2010) have developed a line segmentation technique for handwritten Hindi text. Lajish and Kopparapu (2010) have described a technique for online handwrit-

ten Devanagari script recognition. They have extracted fuzzy directional features for writer independent Devanagari character recognition. Marathi is an Indo-Aryan language spoken in the Indian state of Maharashtra and neighbouring states. Ajmire and Warkhede (2010) have presented a technique based on invariant moments of isolated handwritten Marathi character recognition. The proposed technique is size independent. Shelke and Apte (2011) have presented a multi-stage handwritten character recognition system for Marathi script. They have achieved the recognition accuracy of 96.1 and 94.2% respectively, for training and testing data sets with wavelet approximation features. They have also achieved 98.7 and 96.2% recognition accuracy, respectively, for training and testing samples with modified wavelet features. Belhe et al. (2012) have presented a Hindi handwritten word recognition system. They have used HMM and tree classifier for recognition and obtained a recognition accuracy of 89.0% using 10,000 Hindi words.

5.3 Gujarati

Antani and Agnihotri (1999) are pioneers in attempting Gujarati printed text recognition. For experimental results, they have used dataset of scanned images of printed Gujarati texts collected from various internet sites. Dholakia et al. (2007) attempted to use wavelet features and k-NN classifier on the printed Gujarati text recognition system. They have achieved a recognition accuracy of 96.7% with k-NN classifier. Prasad et al. (2009) have furnished a unique technique called pattern matching for Gujarati script recognition. In this technique, they have identified a character by its shape. Gohell et al. (2015) have presented a low level stroke feature based method for recognition of online handwritten Gujarati characters and numerals. They have accomplished a recognition accuracy of 95, 93 and 90% for numerals dataset, characters dataset and combine dataset of numerals and characters, respectively. Ardeshana et al. (2016) have extracted DCT features for handwritten Gujarati character recognition. They have achieved a recognition accuracy of 78.05% for 22,000 samples using Naïve Bayes classifier. Patel and Kayasth (2017) have presented a recognition system for offline handwritten Gujarati numerals. They have extracted various features namely, hole, straight-line, number of open/end edge and open edge present in different zone for recognition.

5.4 Gurmukhi

Gurmukhi script is used for writing the Punjabi language. Lehal and Singh (1999) have presented a hybrid classification scheme for printed Gurmukhi script recognition. Using this scheme, they have achieved a recognition accuracy of 91.6%. A post processor for Gurmukhi script has been proposed by Lehal et al. (2001). Jindal et al. (2005) have proposed a solution for touching character segmentation of printed Gurmukhi script. Also, they have provided a very useful solution for overlapping lines segmentation in various Indian scripts (2007). They have proposed a technique for segmentation of degraded Gurmukhi script word into upper, middle and lower zones. They have provided a degraded printed Gurmukhi script recognition system. Sharma et al. (2008) have used elastic matching technique for online handwritten Gurmukhi script recognition. Sharma et al. (2009) have expounded a method to rectify the recognition results of handwritten and machine printed Gurmukhi OCR systems. Sharma and Lehal (2009) have set in an algorithm for removal of the field frame boundary of the hand filled forms in Gurmukhi script. Sharma and Jhajj (2010) have extracted zoning features for handwritten Gurmukhi character recognition. They have employed two classifiers, namely, k-NN and SVM. They have achieved maximum recognition accuracy of 72.5 and 72.0%, respectively with k-NN and SVM. Kumar et al. (2013) have presented a novel feature extraction technique

for offline handwritten Gurmukhi character recognition. They have also presented efficient feature extraction techniques based on curvature features for offline handwritten Gurmukhi character recognition (2014a). Kumar et al. (2013) have also presented a character recognition using principal component analysis. They have explored *k*-NN and SVM classifiers for offline handwritten character recognition (Kumar et al. 2011a, b, c, d, 2012).

5.5 Kannada

Kannada is part of the most widely used scripts of Southern India and is spoken by more than fifty million people in India. A little work has been conducted for handwritten Kannada script recognition. Ashwin and Sastry (2002) have presented a font and size independent OCR system for printed Kannada documents. They extracted features based on the foreground pixels in the radial and angular directions. They achieved maximum recognition accuracy of 94.9% using SVM classifier. Sharma et al. (2006) have employed a quadratic classifier for offline handwritten Kannada numerals recognition. They have achieved maximum recognition accuracy of 98.5% using a fivefold cross validation technique. Kunte and Samuel (2007) have presented efficient printed Kannada text recognition system. They considered invariant moments and Zernike moments as features and Neural Network (NN) as classifier. They obtained a recognition accuracy of 96.8% using 2500 characters. Acharya et al. (2008) have come up with a handwritten Kannada numerals recognition system. They have used structural features and multilevel classifiers for recognition. Rajashekararadhya and Ranjan (2008a) have evolved a technique based on zoning and distance metric features. They have utilized feed forward back propagation neural network and obtained recognition accuracy of about 98.0% for Kannada numerals. They have also achieved a recognition accuracy of 97.8% for Kannada numerals with zoning and distance metric features and SVM classifier (2008b). They have utilized Nearest Neighbour classifier for recognition and obtained 97.8% recognition rate for Kannada numerals (2009a). Rajashekararadhya and Ranjan (2009b) have extracted zoning features for offline handwritten numerals of four widely used Indian scripts. For Kannada numerals, they have obtained a recognition accuracy of 98.7% with SVM classifier. Rampalli and Ramakrishnan (2011) have presented an online handwritten Kannada character recognition system which works in combination with an offline handwriting recognition system. They improved the accuracy of online handwriting recognizer by 11% when its combination with offline handwriting recognition system is used. Venkatesh and Ramakrishnan (2011) have presented a technique for fast recognition of online handwritten Kannada characters. Using this technique, they obtained an average accuracy of 92.6% for Kannada characters. Ramakrishnan and Shashidhar (2013) have addressed the challenges in segmentation of online handwritten isolated Kannada words. They achieved 94.3% segmentation accuracy using attention feed-based segmentation technique. Pasha and Padma (2015) have discussed about wavelet transforms and structural features for handwritten Kannada character recognition. They have achieved a recognition accuracy of 91.0 and 97.6% for characters and numerals, respectively. Karthik and Srikanta (2016) have presented a novel approach for handwritten Kannada text recognition using a combination of histogram of gradient features and SVM classifier.

5.6 Malayalam

Malayalam is one of the popular scripts of Southern India. It is the eighth most widely used script in India. Lajish (2007) has presented a system based on fuzzy zoning and normalized vector distance measures for recognition of offline handwritten Malayalam characters.

He has also presented a method for offline handwritten segmented Malayalam character recognition (2008). John et al. (2007) have presented a method based on wavelet transform for offline handwritten Malayalam character recognition. Rajashekararadhya and Ranjan (2008b) have developed a technique of feature extraction for Malayalam script recognition. They have also obtained a recognition accuracy of 96.5% with SVM for Malayalam numerals recognition (2009a). Arora and Namboodiri (2010) have proposed a system for online handwritten Malayalam character recognition. The system achieves stroke level accuracy of 97.9%. Rahiman et al. (2010) have evolved an algorithm which accepts the scanned image of handwritten characters as input and produces editable Malayalam characters in a predefined format as output. Sreeraj and Idicula (2010) have presented a technique for online handwritten Malayalam character recognition. They have employed the k-NN classifier and achieved a recognition accuracy of 98.1%. Sunija et al. (2016) have presented a comparative study of various classifiers for Malayalam dialect recognition system. They have analyzed that a recognition accuracy of 90.2, 88.2 and 84.1% has been accomplished using ANN, SVM, and Naïve Bayes classifier, respectively. Baiju and Sabeerath (2016) have compared K-NN, MLP, and SVM classifiers for online handwritten Malayalam text recognition. They have achieved maximum recognition accuracy of 95.12% for Malayalam character recognition using SVM classifier with RBF kernel.

5.7 Oriya

The Oriya OCR system has been developed at the Indian Statistical Institute, Kolkata by Pal and Chaudhuri (1997). They have utilized the Hough transform based technique for skew angle detection for Oriya alphabets recognition. Tripathy and Pal (2004) have segmented Oriya handwritten text using water reservoir based technique. Roy et al. (2005) dealt with offline unconstrained handwritten Oriya numerals recognition. They have achieved a recognition accuracy of 90.4% using NN classifier with a rejection rate of about 1.84%. Bhowmik et al. (2006) have presented HMM based Oriya numerals recognition system and they have achieved a recognition accuracy of 95.9 and 90.6% for training and testing sets, respectively. Pal et al. (2007d) have used curvature features for Oriya numerals recognition. They have obtained a recognition accuracy of 94.6% using this system. Raj (2015) has presented an optical character recognition for Oriya script. They have considered structural features and a novel combination of a binary tree and Naïve Bayes classifier for recognition purpose. Chaudhary et al. (2015) have used histogram of gradient features and ANN classifier for recognition of printed Oriya characters. Bhoi et al. (2015) have presented Oriya handwritten text recognition system using Hidden Markov Model (HMM). They have extracted concavity features for recognition.

5.8 Tamil

Aparna et al. (2004) have presented a system for online handwritten Tamil character recognition. They have used shape based features including dot, line terminal, bumps and cusp in their work. Deepu et al. (2004) have presented an online handwritten Tamil character recognition using PCA. Prasanth et al. (2007) have described a character based elastic matching technique for online handwritten Tamil character recognition. Bharath and Madhvanath (2011) have used HMM for Tamil word recognition system. They have achieved maximum recognition accuracy of 98.0%. Sundaram and Ramakrishnan (2013) have proposed script-dependent approach to segment online handwritten isolated Tamil words into its constituent symbols. They tested their proposed scheme on a set of 10,000 isolated handwritten words. Sundaram

and Ramakrishnan (2014) reduced the error rate of the Tamil symbol recognition system by reevaluating certain decisions of the SVM classifier. Rajashekararadhya and Ranjan (2008b) have come up with a zoning based feature extraction technique for recognition of offline handwritten numerals of four widely used Indian scripts. They have utilized nearest neighbour, feed forward back propagation neural network and SVM classifiers for recognition. They have obtained a recognition accuracy of 96.1% for Tamil character recognition. Janani et al. (2016) have recognized and analyzed of Tamil inscriptions and mapping of image processing techniques. Elakkiya et al. (2017) have presented Tamil text recognition using K-NN classifier. They have achieved a recognition accuracy of 91.0%.

5.9 Telugu

Prasanth et al. (2007) have used elastic matching technique for online handwritten Telugu character recognition. They have obtained a recognition accuracy of 90.6%. Pal et al. (2007b) have used direction information for Telugu character recognition. They have used a fivefold cross validation technique and obtained a recognition accuracy of 99.4% for Telugu character recognition. Rajashekararadhya and Ranjan (2008a) have recognized handwritten Telugu numerals with zoning and distance metric based features. For recognition, they have used feed forward back propagation neural network classifier and obtained a recognition accuracy of 96.0%. Rajashekararadhya and Ranjan (2008b) have proposed an algorithm based on zoning features for offline handwritten numerals recognition of four widely used Indian scripts. They have obtained a recognition accuracy of 98.6% for handwritten Telugu numerals recognition with an SVM classifier. Arora and Namboodiri (2010) have proposed a system for online handwritten Telugu character recognition. They have achieved a stroke level accuracy of 95.1% for Telugu character recognition. Sastry et al. (2014) have extracted zoning based features for Telugu handwritten character recognition. They have achieved a recognition accuracy of 78.0% using zoning based features. Jyothi et al. (2015) have presented innovative feature sets for Telugu character recognition. They have considered Discrete Wavelet Transformation (DWT), Projection Profile (PP) and Singular Value Decomposition (SVD) features and for classification they have explored k-NN and SVM classifiers. Kinjarapu et al. (2016) have presented an online handwriting recognition system for Telugu script. They have accomplished a recognition accuracy of 90% using a combination of strokes. Prasad and Kanduri (2016) have used zoning based features and Genetic Algorithm for Telugu handwritten character recognition. They have considered k-NN classifier for recognition purpose and achieved a recognition accuracy of 88.8%.

6 Recognition results of non-Indic and Indic scripts

This section presents a brief report on the recognition accuracies achieved by researchers for character and numeral recognition. We have presented their results in Tables 1 and 2 for numerals, Tables 3 and 4 for non-Indic scripts and Tables 5 and 6 for Indic scripts. As illustrated in Table 1, one may note that a recognition accuracy of 99.6% has been achieved for handwritten numerals by Pal et al. (2007b). Feature wise comparative study of numeral recognition is depicted in Table 2. In Table 3, the results of non-Indian scripts are presented. As shown in this table, one may note that a recognition accuracy of 99.4, 99.9, 92.8 and 99.2%, has been achieved for Arabic, French, Japanese and Roman scripts, respectively. Srihari and Leedham (2003) have also presented a good survey on computer methods in forensic handwritten document examination. They have presented various software systems

Table 1 Recognition results of numerals

Authors	Test data size	Feature extraction technique	Classifier	Accuracy (%)
Bhattacharya and Chaudhuri (2003)	5000	Wavelet	MLP	97.2
Bhattacharya et al. (2004)	5000	Wavelet	MLP	98.0
Roy et al. (2004)	12,410	Structural, topological	NN	94.2
Roy et al. (2005)	3850	Directional	Quadratic	94.8
Kunte and Samuel (2007)	11,500	Wavelet	MLP	92.3
Pal and Dutta (2003)	12,000	Water reservoir	Binary tree	92.8
Rajput and Hangarge (2007)	1250	Image fusion	NN	91.0
Pal et al. (2007b)	5638	Directional	MQC	99.6
Lu et al. (2008)	16,000	Directional and density	SOM	97.3
Desai (2010)	3260	Profiles	FFNN	81.7
Purkait and Chanda (2010)	23,392	Morphological	MLP	97.8

Table 2 Feature wise recognition results of numerals

Features	Authors	Test data size	Classifier	Accuracy (%)
Statistical features	Bhattacharya and Chaudhuri (2003)	5000	MLP	97.2
	Bhattacharya et al. (2004)	5000	MLP	98.0
	Roy et al. (2005)	3850	Quadratic	94.8
	Pal et al. (2007b)	5638	MQC	99.6
	Kunte and Samuel (2007)	11,500	MLP	92.3
	Desai (2010)	3260	FFNN	81.7
Structural features	Pal and Dutta (2003)	12,000	Binary tree	92.8
	Roy et al. (2004)	12,410	NN	94.2
	Rajput and Hangarge (2007)	1250	NN	91.0
	Purkait and Chanda (2010)	23,392	MLP	97.8

that automate some of the examination processes and have included verification methods to provide the degree of match between a questioned and known document. Feature wise comparative study for non-Indic scripts recognition is presented in Table 4. In Table 5, the results on Indic scripts have been presented. It can be seen that a lot of work has been done on Bangla, Devanagari and Kannada scripts. Some work has also been done to recognize the Gurmukhi, Malayalam, Oriya and Tamil scripts as given in this table. As depicted in Table 5, for the Bangla script, maximum recognition accuracy of 97.6% has been achieved by Roy et al. (2005). In Devanagari script, maximum recognition accuracy of 99.0% has been achieved by Pal et al. (2007b). They have used directional features and MQDF classifier for recognition. Kunte and Samuel (2007) have achieved a maximum recognition accuracy of 96.8% for Kannada characters. They have tested their technique with 1000 samples of 50-class problem. For all classes of Kannada script, maximum recognition accuracy of 92.6% has been achieved by Venkatesh and Ramakrishnan (2011). They have considered 26,926 samples for testing data set. Arora and Namboodiri (2010) have achieved a recognition accuracy of 95.8% for Malayalam character recognition. They have tested their technique with 7348 samples of Malayalam characters. Joshi et al. (2004) have achieved a maximum recognition accuracy of 91.5% for Tamil character recognition. They have considered 4860 samples of 156 classes for

Table 3 Recognition results of non-Indic scripts

Authors	Script (*Language*)	Data set	Text mode	Feature extraction technique	Classifier	Accuracy (%)
Alaei et al. (2009)	Arabic (*Arabian*)	20,000	Handwritten	Chain code direction	SVM	99.4
Tran et al. (2010)	French (*French*)	400	Handwritten	Statistical and structural	SVM	99.9
Chanda et al. (2007a)	Japanese (*Japanese*)	11,304	Printed	Structural, topological	Tree	98.8
Zhu et al. (2010)	Japanese (*Japanese*)	35,686	Handwritten	Geometric features	SVM	92.8
Alaei et al. (2010b)	Arabic (*Persian*)	20,000	Handwritten	Modified chain code direction	SVM	98.1
Chanda et al. (2007a)	Roman (*English*)	11,304	Printed	Structural, topological	SVM	99.4
Pal and Chaudhuri (1997)	Roman (*English*)	11,875	Handwritten	Chain code	MQDF	99.0
Roy et al. (2005)	Roman (*English*)	5000	Handwritten	Fractal dimension, topological	MLP, SVM, k-NN, MQDF	99.2
Jayadevan et al. (2010)	Roman (*English*)	5400	Handwritten	Directional	MQDF	97.0
Chanda et al. (2007b)	Thai (*Thai*)	11,304	Handwritten	Structural, topological, water reservoir concept	SVM	99.4

testing data set. For offline handwritten Gurmukhi script, a recognition accuracy of 91.8% has been achieved by Kumar et al. (2014b). They have tested their technique with 5600 samples of 56-class problem. Nonetheless till now, there is no complete recognition system available for recognition of Indic script. In Table 6, we have presented comparisons of recognition results using the same feature types with different dataset and classifiers.

7 Suggestions on future directions

In optical character recognition field, a lot of directions are possible for future research as proposed algorithms used for segmentation task can be extended further for improving the recognition accuracy because segmentation is essential part of document recognition process. The following are some suggestions on future research directions in character and numeral recognition:

a. There must be multiple standard handwritten character databases for non-Indic scripts and database should be adequately large in size.
b. New features can be proposed to improve the recognition accuracy of different scripts. There is a need to develop the standard database for Devanagari, Gurmukhi scripts etc.

Table 4 Feature wise comparative study of recognition results for non-Indic scripts

Features	Authors	Script	Test data size	Classifier	Accuracy (%)
Structural features	Chanda et al. (2007a)	Japanese (*Japanese*)	11,304	Tree	98.8
	Chanda et al. (2007a)	Roman (*English*)	11,304	SVM	99.4
	Chanda et al. (2007b)	Thai (*Thai*)	11,304	SVM	99.4
	Tran et al. (2010)	French (*French*)	400	SVM	99.9
Statistical features	Alaei et al. (2009)	Arabic (*Arabian*)	20,000	SVM	99.4
	Alaei et al. (2010b)	Arabic (*Persian*)	20,000	SVM	98.1
	Zhu et al. (2010)	Japanese (*Japanese*)	35,686	MLP	97.8
	Pal and Chaudhuri (1997)	Roman (*English*)	11,875	MQDF	99.0
	Roy et al. (2005)	Roman (*English*)	5000	MLP, SVM, k-NN, MQDF	99.2

c. A combination of statistical and structural features should be considered for extracting the relevant information about characters.

d. More research should focus on the image transformation based representations

e. More research can be carried out on the feature selection techniques and classification techniques for different scripts recognition. An optical character recognition system could be developed for multi-font style characters.

f. Most of the work reported on fair quality documents. Sophisticated studies on degraded documents are not undertaken by the scientists in the development character recognition system.

g. Experiments should be made to observe the effect of degraded quality paper as well as noise of various types, and take corrective measures.

8 Conclusions

In this paper, we have surveyed the character and numeral recognition work that has been done on non-Indic and Indic scripts. We have assessed the work done for various Indic scripts, i.e., Bangla, Devanagari, Gujarati, Gurmukhi, Kannada, Malayalam, Oriya, Tamil and Telugu. Also, we have presented the work done for recognition of various non-Indic scripts, i.e., Arabic, French, Japanese, Roman, and Thai. We have presented recognition accuracies achieved for character and numeral recognition of different non-Indic and Indic scripts. We have seen that the efficient techniques used for non-Indic scripts may be used for Indic scripts (printed text and handwritten text) so that accuracy of recognition may be increased as non-Indic scripts. One of the key inspirations of early development of character and numeral recognition system was a reading help for the visually handicapped. One possible way of achieving this goal is to convert the character and numeral output into speech format.

Table 5 Recognition results of Indic scripts

Authors	Script (*Language*)	Data set	Text mode	Feature extraction technique	Classifier	Accuracy
Reddy et al. (2012a)	Bangla (*Assamese*)	207	Handwritten	Projection profiles, DCT, chain code histograms	Vector quantization	96.6% (online) and 97.6% (offline)
Reddy et al. (2012b)	Bangla (*Assamese*)	18,000	Handwritten	Derivative of *x* and *y*	HMM classifier	97.1%
Sarma et al. (2013)	Bangla (*Assamese*)	200	Handwritten	Derivative of *x* and *y*	HMM and SVM	96.5% (HMM) and 96.8% (SVM)
Pal and Chaudhuri (1994)	Bangla (*Bengali*)	20,000	Printed	Statistical	Tree classifier	96.0%
Garain et al. (2002)	Bangla (*Bengali*)	2440	Handwritten	Pen trajectory	Template matching	96.3%
Roy et al. (2005)	Bangla (*Bengali*)	4752	Handwritten	Water reservoir, fractal dimensions	Neural network	97.6%
Pal et al. (2006)	Bangla (*Bengali*)	2692	Handwritten	Directional	MQDF	87.2%
Pal and Chaudhuri (2001)	Devanagari (*Hindi*)	5000	Printed	Structural and statistical	Tree	98.3%
Bajaj et al. (2002)	Devanagari (*Hindi*)	2460	Handwritten	Moments and descriptive component	Multiple connectionist	89.7%
Joshi et al. (2005)	Devanagari (*Hindi*)	1487	Handwritten	Gaussian low pass filters	Feature based	94.5%
Hanmandlu et al. (2007)	Devanagari (*Hindi*)	4750	Handwritten	Normalized distance	Fuzzy set	90.7%
Pal et al. (2007b)	Devanagari (*Hindi*)	5424	Handwritten	Directional	MQDF	99.0%

Table 5 continued

Authors	Script (*Language*)	Data set	Text mode	Feature extraction technique	Classifier	Accuracy
Pal et al. (2007c)	Devanagari (*Hindi*)	5424	Handwritten	Curvature and gradient	SVM and *k*-NN	94.9%
Chanda et al. (2009)	Devanagari (*Hindi*)	3949	Handwritten	Dimensional and gradient	SVM	98.5%
Lehal and Singh (1999)	Gurmukhi (*Punjabi*)	3000	Printed	Sub character	Binary decision tree and NN	96.6%
Lehal et al. (2001)	Gurmukhi (*Punjabi*)	30,000	Printed	zoning, local features and global features	Binary decision tree and NN	97.0%
Sharma et al. (2008)	Gurmukhi (*Punjabi*)	2460	Online Handwritten	Elastic matching	*k*-Means	87.4%
Jindal et al. (2007)	Gurmukhi (*Punjabi*)	2855	Printed	Structural features	SVM	92.5%
Sharma and Jhajj (2010)	Gurmukhi (*Punjabi*)	2050	Handwritten	Zoning	SVM	72.0%
Kumar et al. (2014b)	Gurmukhi (*Punjabi*)	5600	Handwritten	Hierarchical technique	SVM	91.8%
Pal et al. (2007b)	Kannada (*Kannada*)	4820	Handwritten	Directional	Modified quadratic	98.7%
Rajashekararadhya and Ranjan (2009a)	Kannada (*Kannada*)	4000	Handwritten	Zone and distance metric	Feed forward BPN network	98.7%
Rampalli and Ramakrishnan (2011)	Kannada	6195	Handwritten	Directional distance distribution, transitions, projection profiles	SVM	89.7%
Venkatesh and Ramakrishnan (2011)	Kannada	26,926	Handwritten	Quantized slope, quartile features	DTW	92.6%
Lajish (2007)	Malayalam (*Malayalam*)	15,752	Handwritten	Fuzzy zoning	Class modular NN	78.9%
John et al. (2007)	Malayalam (*Malayalam*)	4950	Handwritten	Wavelet transform	MLP	73.8%

Table 5 continued

Authors	Script (Language)	Data set	Text mode	Feature extraction technique	Classifier	Accuracy
Rajashekararadhya and Ranjan (2009a)	Malayalam (Malayalam)	2000	Handwritten	Zoning	NN, feed forward NN	96.5%
Arora and Namboodiri (2010)	Malayalam (Malayalam)	7348	Handwritten	Strokes	HMM, DTW	95.8%
Sreeraj and Idicula (2010)	Malayalam (Malayalam)	2591	Handwritten	Time domain and direction	k-NN	98.1%
Roy et al. (2005)	Oriya (Oriya)		Handwritten	Chain code histogram	Neural network	90.4%
Bhowmik et al. (2006)	Oriya (Oriya)	5970	Handwritten	Strokes	HMM	95.9%
Roy and Pal (2006)	Oriya (Oriya)	2500	Handwritten	Water reservoir, fractal, topological features	Neural network	97.7%
Pal et al. (2007b)	Oriya (Oriya)	5638	Handwritten	curvature features	Quadratic	94.6%
Deepu et al. (2004)	Tamil (Tamil)	21,840	Handwritten	Gaussian low pass filters	PCA and NN	95.3%
Pal et al. (2007b)	Tamil (Tamil)	2690	Handwritten	Directional	Modified quadratic	98.5%
Prasanth et al. (2007)	Tamil (Tamil)	4680	Handwritten	Shape context	NN with DTW	89.8%
Rajashekararadhya and Ranjan (2009a)	Tamil (Tamil)	2000	Handwritten	Zoning	NN, feed forward PNN and SVM	96.1%
Mondal et al. (2010)	Tamil (Tamil)	26,926	Handwritten	Direction code histogram	NN, MLP and HMM	98.0%
Sundaram and Ramakrishnan (2008)	Tamil (Tamil)	1560	Handwritten	Local features	Modified mahalanobis distance measure	83.4%
Prasanth et al. (2007)	Telugu (Telugu)	156	Handwritten	Shape context	NN with DTW	90.6%
Rajashekararadhya and Ranjan (2009a)	Telugu (Telugu)	2000	Handwritten	Zone and distance metric	Feed forward BPNN	98.6%
Sastry et al. (2010)	Telugu	Not mentioned	Handwritten	3D	Decision tree	93.1%

Table 6 Feature wise comparative study of recognition results for Indic scripts

Features	Authors	Script	Test data size	Classifier	Accuracy
Structural features	Chanda et al. (2009)	Devanagari (*Hindi*)	3949	SVM	98.5%
	Garain et al. (2002)	Bangla (*Bengali*)	2440	Template matching	96.3%
	Jindal et al. (2007)	Gurmukhi (*Punjabi*)	2855	SVM	92.5%
	Joshi et al. (2005)	Devanagari (*Hindi*)	1487	Feature based	94.5%
	Lehal and Singh (1999)	Gurmukhi (*Punjabi*)	3000	Binary decision tree and NN	96.6%
	Pal and Chaudhuri (2001)	Devanagari (*Hindi*)	5000	Tree	98.3%
	Prasanth et al. (2007)	Tamil (*Tamil*)	4680	NN with DTW	89.8%
	Prasanth et al. (2007)	Telugu (*Telugu*)	156	NN with DTW	90.6%
	Venkatesh and Ramakrishnan (2011)	Kannada	26,926	DTW	92.6%
Statistical features	Hanmandlu et al. (2007)	Devanagari (*Hindi*)	4750	Fuzzy set	90.7%
	Kumar et al. (2014b)	Gurmukhi (*Punjabi*)	5600	SVM	91.8%
	Kumar et al. (2016)	Gurmukhi (*Punjabi*)	10,500	SVM	95.8%
	Lajish (2007)	Malayalam (*Malayalam*)	15,752	Class modular NN	78.9%
	Pal et al. (2007c)	Devanagari (*Hindi*)	5424	SVM and *k*-NN	94.9%
	Pal et al. (2007b)	Kannada (*Kannada*)	4820	Modified quadratic	98.7%
	Rajashekararadhya and Ranjan (2009a)	Kannada (*Kannada*)	4000	Feed forward BPN network	98.7%
	Rajashekararadhya and Ranjan (2009a)	Malayalam (*Malayalam*)	2000	NN, feed forward NN	96.5%
	Rajashekararadhya and Ranjan (2009a)	Tamil (*Tamil*)	2000	NN, feed forward PNN and SVM	96.1%

Table 6 continued

Features	Authors	Script	Test data size	Classifier	Accuracy
	Rajashekararadhya and Ranjan (2009a)	Telugu (*Telugu*)	2000	Feed forward BPNN	98.6%
	Sharma and Jhajj (2010)	Gurmukhi (*Punjabi*)	2050	SVM	72.0%
	Pal and Chaudhuri (1994)	Bangla (*Bengali*)	20,000	Tree classifier	96.0%
Parametric features	Reddy et al. (2012b)	Bangla (*Assamese*)	18,000	HMM classifier	97.1%
	Sarma et al. (2013)	Bangla (*Assamese*)	200	HMM and SVM	96.5% (HMM) and 96.8% (SVM)
Projection profiles and local features	Bajaj et al. (2002)	Devanagari (*Hindi*)	2460	Multiple connectionist	89.7%
	Lehal et al. (2001)	Gurmukhi (*Punjabi*)	30,000	Binary decision tree and NN	97.0%
	Rampalli and Ramakrishnan (2011)	Kannada	6195	SVM	89.7%
	Reddy et al. (2012a)	Bangla (*Assamese*)	207	Vector quantization	96.6% (online) and 97.6% (offline)
	Roy et al. (2005)	Bangla (*Bengali*)	4752	Neural network	97.6%
	Sharma et al. (2008)	Gurmukhi (*Punjabi*)	2460	*k*-Means	87.4%
	Sundaram and Ramakrishnan (2008)	Tamil (*Tamil*)	1560	Modified mahalanobis distance measure	83.4%
Stroke information for online handwriting	Arora and Namboodiri (2010)	Malayalam (*Malayalam*)	7348	HMM, DTW	95.8%
	Bhowmik et al. (2006)	Oriya (*Oriya*)	5970	HMM	95.9%

References

Abd MA, Paschos G (2007) Effective Arabic character recognition using support vector machines. Innov Adv Tech Comput Inf Sci Eng 7–11

Acharya D, Reddy NVS, Makkithaya K (2008) Multilevel classifiers in recognition of handwritten Kannada numerals. Proc World Acad Sci Eng Technol (WASET) 18:274–279

Afroge S, Ahmed B, Mahmud F (2016) Optical character recognition using back propagation neural network. In: Proceedings of the 2nd international conference on electrical, computer and telecommunication engineering (ICECTE), pp 1–4

Ajmire PE, Warkhede SE (2010) Handwritten Marathi character (vowel) recognition. Adv Inf Min 2(2):11–13

Alaei A, Pal U, Nagabhushan P (2009) Using modified contour features and SVM based classifier for the recognition of Persian/Arabic handwritten numerals. In: Proceedings of the 7th international conference on advances in pattern recognition (ICAPR), pp 391–394

Alaei A, Nagabhushan P, Pal U (2010a) A baseline dependent approach for Persian handwritten character segmentation. In: Proceedings of the 20th international conference on pattern recognition (ICPR), pp 1977–1980

Alaei A, Nagabhushan P, Pal U (2010b) A new two-stage scheme for the recognition of Persian handwritten characters. In: Proceedings of the 12th international conference on frontiers in handwriting recognition (ICFHR), pp 130–135

Almuallim H, Yamaguchi S (1987) A method of recognition of Arabic cursive handwriting. IEEE Trans Pattern Anal Mach Intell 9(5):715–722

Althobaiti H, Lu C (2017) A survey on Arabic optical character recognition and an isolated handwritten Arabic character recognition algorithm using encoded freeman chain code. In: Proceedings of the 51st annual conference on information sciences and systems

Antani S, Agnihotri L (1999) Gujarati character recognition. In: Proceedings of the 5th international conference on document analysis and recognition (ICDAR), pp 418–421

Aparna KH, Subramanian V, Kasirajan M, Prakash GV, Chakravarthy VS, Madhvanath S (2004) Online handwriting recognition for Tamil. In: Proceedings of the 9th international workshop on frontiers in handwriting recognition (IWFHR), pp 438–443

Ardeshana M, Sharma AK, Adhyaru DM, Zaveri TH (2016) Handwritten Gujarati character recognition based on discrete cosine transform. In: Proceedings of the IRF-IEEE forum international conference, pp 23–26

Arora A, Namboodiri AM (2010) A hybrid model for recognition of online handwriting in Indian scripts. In: Proceedings of the 12th international conference on frontiers in handwriting recognition (ICFHR), pp 433–438

Asavareongchai N, Giarta E (2016) Recognition of Thai characters and text from document templates. Project report, pp 1–7

Ashwin TV, Sastry PS (2002) A font and size-independent OCR system for printed Kannada documents using support vector machines. SADHANA 27(1):35–58

Baiju KB, Sabeerath K (2016) Online recognition of Malayalam handwritten scripts—a comparison using KNN, MLP and SVM. In: Proceedings of the international conference on advances in computing, communications and informatics, pp 2078–2083

Bajaj R, Dey L, Chaudhury S (2002) Devanagari numeral recognition by combining decision of multiple connectionist classifiers. SADHANA 27(1):59–72

Bansal V, Sinha RMK (2000) Integrating knowledge sources in Devanagari text recognition system. IEEE Trans Syst Man Cybern Part A 30(4):500–505

Belhe S, Paulzagade C, Deshmukh A, Jetley S, Mehrotra K (2012) Hindi handwritten word recognition using HMM and symbol tree. In: Proceedings of the workshop on document analysis and recognition (DAR), pp 9–14

Bharath A, Madhvanath S (2011) HMM-based lexicon-driven and lexicon-free word recognition for online handwritten Indic scripts. IEEE Trans Pattern Anal Mach Intell (PAMI) 34(4):670–682

Bhattacharya U, Chaudhuri BB (2003) A majority voting scheme for multi-resolution recognition of hand-printed numerals. In: Proceedings of the 7th international conference on document analysis and recognition (ICDAR), pp 16–20

Bhattacharya U, Vajda S, Mallick A, Chaudhuri BB, Belaid A (2004) On the choice of training set, architecture and combination rule of multiple MLP classifiers for multi-resolution recognition of handwritten characters. In: Proceedings of the 9th international workshop on frontiers in handwriting recognition (IWFHR), pp 419–424

Bhattacharya U, Shridhar M, Parui SK (2006) On recognition of handwritten Bangla characters. In: Proceedings of the international conference on computer vision, graphics and image processing (ICVGIP), pp 817–828

Bhattacharya U, Gupta BK, Parui SK (2007) Direction code based features for recognition of online handwritten characters of Bangla. In: Proceedings of the 9th international conference on document analysis and recognition (ICDAR), vol 1, pp 58–62

Bhoi S, Dogra DP, Roy PP (2015) Handwritten text recognition in Odia script using hidden Markov model. In: Proceedings of the 5th national conference on computer vision, pattern recognition, image processing and graphics, pp 1–3

Bhowmik TK, Parui SK, Bhattacharya U, Shaw B (2006) An HMM based recognition scheme for handwritten Oriya numerals. In: Proceedings of the 9th international conference on information technology (ICIT), pp 105–110

Bishnu A, Chaudhuri BB (1999) Segmentation of Bangla handwritten text into characters by recursive contour following. In: Proceedings of the 5th international conference on document analysis and recognition (ICDAR), pp 402–405

Bluche T, Messina R (2016) Faster segmentation-free handwritten Chinese text recognition with character decompositions. In: Proceedings of the 15th international conference on frontiers in handwriting recognition, pp 530–535

Borovikov E (2004) A survey of modern optical character recognition techniques. Project report, pp 1–38

Bunke H, Varga T (2007) Off-line Roman cursive handwriting recognition. Adv Pattern Recognit 165–183

Chanda S, Pal U, Kimura F (2007a) Identification of Japanese and English script from a single document page. In: Proceedings of the 7th IEEE international conference on computer and information technology, pp 656–661

Chanda S, Terrades OR, Pal U (2007b) SVM based scheme for Thai and English script identification. In: Proceedings of the 9th international conference on document analysis and recognition (ICDAR), vol 1, pp 551–555

Chanda S, Pal S, Franke K, Pal U (2009) Two-stage approach for word-wise script identification. In: Proceedings of the 10th international conference on document analysis and recognition (ICDAR), pp 926–930

Chaudhary S, Sharma S, Kumar B (2015) Recognition of printed Oriya script using gradient based features. In: Proceedings of the 2015 annual IEEE India conference, pp 1–5

Das S, Banerjee S (2015) An algorithm for Japanese character recognition. Int J Image Gr Signal Process 1:9–15

Deepu V, Madhvanath S, Ramakrishnan AG (2004) Principal component analysis for online handwritten character recognition. In: Proceedings of the 17th international conference on pattern recognition (ICPR), vol 2, pp 327–330

Desai AA (2010) Gujarati handwritten numeral optical character reorganization through neural network. Pattern Recognit 43(7):2582–2589

Dholakia J, Yajnik A, Negi A (2007) Wavelet feature based confusion character sets for Gujarati script. In: Proceedings of the international conference on computational intelligence and multimedia applications (CIMA), pp 366–370

Dutta A, Chaudhury S (1993) Bengali alpha-numeric character recognition using curvature features. Pattern Recognit 26(12):1757–1770

Elakkiya V, Muthumani I, Jegajothi M (2017) Tamil text recognition using KNN classifier. Adv Nat Appl Sci 11(7):41–45

Garg NK, Kaur L, Jindal MK (2010) A new method for line segmentation of handwritten Hindi text. In: Proceedings of the 7th international conference on information technology: new generations (ITNG), pp 392–397

Garain U, Chaudhuri BB, Pal TT (2002) Online handwritten Indian script recognition: a human motor function based framework. In: Proceedings of the 16th international conference on pattern recognition (ICPR), vol 3, pp 164–167

Gohell CC, Goswam MM, Prajapate YK (2015) On-line Handwritten Gujarati character recognition using low level stroke. In: Proceedings of the third international conference on image information processing, pp 130–134

Grosicki E, Abed H (2009) Handwriting recognition competition. In: Proceedings of the 10th international conference on document analysis and recognition (ICDAR), pp 1398–1402

Hanmandlu M, Murthy OVR, Madasu VK (2007) Fuzzy model based recognition of handwritten Hindi characters. In: Proceedings of the 9th biennial conference of the Australian pattern recognition society on digital image computing and techniques and applications, pp 454–461

He S, Hu X (2016) Chinese character recognition in natural scenes. In: Proceedings of the 9th international symposium on computational intelligence and design, pp 124–127

Hussain R, Raza A, Siddiqi I, Khurshid K, Djeddi C (2015) A comprehensive survey of handwritten document benchmarks: structure, usage and evaluation. EURASIP J Image Video Process 46:1–24

Impedovo S, Dimauro G (1990) An interactive system for the selection of handwritten numeral classes. In: Proceedings of the 10th international conference on pattern recognition (ICPR), pp 563–566

Izadi S, Sadri J, Solimanpour F, Suen CY (2006) A review on Persian script and recognition techniques. In: Proceedings of the conference on Arabic and Chinese handwriting, pp 22–35

Janani G, Vishalini V, Kumar PM (2016) Recognition and analysis of Tamil inscriptions and mapping using image processing techniques. In: Second international conference on science technology engineering and management, pp 181–184

Jayadevan R, Pal U, Kimura F (2010) Recognition of words from legal amounts of Indian bank cheques. In: Proceedings of the international conference on frontiers in handwriting recognition (ICFHR), pp 166–171

Jindal MK, Lehal GS, Sharma RK (2005) Segmentation problems and solutions in printed degraded Gurmukhi script. Int J Signal Process 2(4):258–267

Jindal MK, Sharma RK, Lehal GS (2007) Segmentation of horizontally overlapping lines in printed Indian scripts. Int J Comput Intell Res 3(4):277–286

John R, Raju G, Guru DS (2007) 1D wavelet transform of projection profiles for isolated handwritten Malayalam character recognition. Proceedings of the international conference on computational intelligence and multimedia applications (ICCIMA) 2:481–485

Joshi N, Sita G, Ramakrishnan AG, Madhvanath S (2004) Tamil handwriting recognition using subspace and DTW based classifiers. In: Proceedings of the international conference on neural information processing (ICONIP), pp 806–813

Joshi N, Sita G, Ramakrishnan AG, Deepu V, Madhvanath S (2005) Machine recognition of online handwritten Devanagari characters. In: Proceedings of the 8th international conference on document analysis and recognition (ICDAR), vol 2, pp 1156–1160

Jyothi J, Manjusha K, Kumar MA, Soman KP (2015) Innovative feature sets for machine learning based Telugu character recognition. Indian J Sci Technol 8(24):1–7

Kacem A, Aouiti N, Belaid A (2012) Structural features extraction for handwritten Arabic personal names recognition. In: Proceedings of the international conference on frontiers in handwriting recognition (ICFHR), pp 268–273

Karnchanapusakij C, Suwannakat P, Rakprasertsuk W, Dejdumrong N (2009) Online handwriting Thai character recognition. In: Proceedings of the 6th international conference on computer graphics, imaging and visualization (CGIV), pp 323–328

Karthik S, Srikanta MK (2016) Segmentation and recognition of handwritten Kannada text using relevance feedback and histogram of oriented gradients: a novel approach. Int J Adv Comput Sci Appl 7(1):472–476

Kinjarapu AA, Yelavarti KC, Valurouthu KP (2016) Online recognition of handwritten Telugu script characters. In: Proceedings of the international conference on signal processing, communication, power and embedded system, pp 426–432

Kobchaisawat T, Chalidabhongse T (2015) A method for multi-oriented Thai text localization in natural scene images using convolutional neural network. In: Proceedings of the international conference on signal and image processing applications, pp 220–225

Kumar D (2008) AI approach to hand written Devanagri script recognition. In: Proceedings D the IEEE region 10th international conference on EC3-energy, computer, communication and control systems vol 2, pp 229–237

Kumar M, Jindal MK, Sharma RK (2011a) Review on OCR for handwritten Indian scripts character recognition. In: Proceedings of the first international conference on digital image processing and pattern recognition (DPPR), Tirunelveli, Tamil Nadu, vol 205, pp 268–276

Kumar M, Jindal MK, Sharma RK (2011b) k-Nearest neighbor based offline handwritten Gurmukhi character recognition. In: Proceedings of the international conference on image information processing (ICIIP), Jaypee University of Information Technology, Waknaghat (Shimla), pp 1–4

Kumar M, Sharma RK, Jindal MK (2011c) Classification of characters and grading writers in offline handwritten Gurmukhi script. In: Proceedings of the international conference on image information processing (ICIIP), Jaypee University of Information Technology, Waknaghat (Shimla), pp 1–4

Kumar M, Sharma RK, Jindal MK (2011d) SVM based offline handwritten Gurmukhi character recognition. In: Proceedings of the international workshop on soft computing applications and knowledge discovery (SCAKD), National Research University Higher School of Economics, Moscow (Russia), pp 51–62

Kumar M, Jindal MK, Sharma RK (2012) Offline handwritten Gurmukhi character recognition: study of different features and classifiers combinations. In: Proceedings of the workshop on document analysis and recognition (IWDAR), IIT Bombay, pp 94–99

Kumar M, Sharma RK, Jindal MK (2013a) A novel feature extraction technique for offline handwritten Gurmukhi character recognition. IETE J Res 59(6):687–692

Kumar M, Jindal MK, Sharma RK (2013b) PCA based offline handwritten Gurmukhi character recognition. Smart Comput Rev 3(5):346–357

Kumar M, Sharma RK, Jindal MK (2014a) Efficient feature extraction techniques for offline handwritten Gurmukhi character recognition. Natl Acad Sci Lett 37(4):381–391

Kumar M, Sharma RK, Jindal MK (2014b) A novel hierarchical technique for offline handwritten Gurmukhi character recognition. Natl Acad Sci Lett 37(6):567–572

Kumar M, Jindal MK, Sharma RK (2016) Offline handwritten Gurmukhi character recognition: analytical study of different transformations. Proc Natl Acad Sci India Sect A Phys Sci 87(1):137–143

Koundal K, Kumar M, Garg NK (2017) Punjabi optical character recognition: a survey. Indian J Sci Technol 10(19):1–8

Kunte RS, Samuel RDS (2007) A simple and efficient optical character recognition system for basic symbols in printed Kannada text. SADHANA 32(5):521–533

Lajish VL (2007) Handwritten character recognition using perceptual fuzzy-zoning and class modular neural networks. In: Proceedings of the 4th international conference on innovations in information technology (ICIIT), pp 188–192

Lajish VL (2008) Handwritten character recognition using gray-scale based state-space parameters and class modular NN. In: Proceedings of the international conference on signal processing, communications and networking (ICSCN), pp 374–379

Lajish VL, Kopparapu SK (2010) Fuzzy directional features for unconstrained on-line Devanagari handwriting recognition. In: Proceedings of the national conference on communications (NCC), pp 1–5

Lehal GS, Singh C (1999) Feature extraction and classification for OCR of Gurmukhi script. Vivek 12(2):2–12

Lehal GS, Singh C, Lehal R (2001) A shape based post processor for Gurmukhi OCR. In: Proceedings of the 6th international conference on document analysis and recognition (ICDAR), pp 1105–1109

Liang J, Zhu B, Kumagai T, Nakagawa M (2016) Character-position-free on-line handwritten Japanese text recognition by two segmentation methods. IEICE Trans Inf Syst E99–D(4):1172–1181

Liao SX, Chiang A, Lu Q, Pawlak M (2002) Chinese character recognition via Gegenbauer moments. In: International conference on pattern recognition, pp 485–488

Liu J, Ma SP (1996) An overview of printed Chinese character recognition techniques. In: Proceedings of the international conference on Chinese computing, Singapore, pp 325–333

Liwicki M, Bunke H (2007) Combining on-line and off-line systems for handwriting recognition. In: Proceedings of the international conference on document analysis and recognition (ICDAR), pp 372–376

Lorigo LM, Govindaraju V (2006) Offline Arabic handwriting recognition: a survey. IEEE Trans Pattern Anal Mach Intell (PAMI) 28(5):712–724

Lu S, Tu X, Lu Y (2008) An improved two-layer SOM classifier for handwritten numeral recognition. In: Proceedings of the international conference on intelligent information technology, pp 367–371

Modi H, Parikh MC (2017) A review on optical character recognition techniques. Int J Comput Appl 160(6):20–24

Mondal T, Bhattacharya U, Parui SK, Das K, Mandalapu D (2010) On-line handwriting recognition of Indian scripts-the first benchmark. In: Proceedings of the 12th international conference on frontiers in handwriting recognition (ICFHR), pp 200–205

Nakagawa M, Zhu B, Onuma M (2005) A model of on-line handwritten Japanese text recognition free from line direction and writing format constraints. IEICE Trans Inf Syst E88–D(8):1815–1822

Pal U, Chaudhuri BB (1994) OCR in Bangla: an Indo-Bangladeshi language. In: Proceedings of the 12th international conference on pattern recognition (ICPR), pp 269–274

Pal U, Chaudhuri BB (1997) Automatic separation of words in multi-lingual multi-script Indian documents. In: Proceedings of the 4th international conference on document analysis and recognition (ICDAR), vol 2, pp 576–579

Pal U, Chaudhuri BB (2001) Automatic identification of English, Chinese, Arabic, Devanagari and Bangla script line. In: Proceedings of the 6th international conference on document analysis and recognition (ICDAR), pp 790–794

Pal U, Dutta S (2003) Segmentation of Bangla unconstrained handwritten text. In: Proceedings of the 7th international conference on document analysis and recognition (ICDAR), pp 1128–1132

Pal U, Jayadevan R, Sharma N (2012) Handwriting recognition in Indian regional scripts: a survey of offline techniques. ACM Trans Asian Lang Inf Process 11(1):1–35

Pal U, Roy K, Kimura F (2006) A lexicon driven method for unconstrained Bangla handwritten word recognition. In: Proceedings of the 10th international workshop on frontiers in handwriting recognition (IWFHR), pp 601–606

Pal U, Wakabayashi T, Kimura F (2007a) Handwritten Bangla compound character recognition using gradient feature. In: Proceedings of the 10th international conference on information technology (ICIT), pp 208–213

Pal U, Sharma N, Wakabayashi T, Kimura F (2007b) Handwritten numeral recognition of six popular Indian scripts. In: Proceedings of the 9th international conference on document analysis and recognition (ICDAR), vol 2, pp 749–753

Pal U, Sharma N, Wakabayashi T, Kimura F (2007c) Off-line handwritten character recognition of Devanagari script. In: Proceedings of the 9th international conference on document analysis and recognition (ICDAR), pp 496–500

Pal U, Wakabayashi T, Kimura F (2007d) A system for off-line Oriya handwritten character recognition using curvature feature. In: Proceedings of the 10th international conference on information technology (ICIT), pp 227–229

Pal U, Roy K, Kimura F (2008) Bangla handwritten pin code string recognition for Indian postal automation. In: Proceedings of the 11th international conference on frontiers in handwriting recognition (ICFHR), pp 290–295

Pal U, Wakabayashi T, Kimura F (2009) Comparative study of Devanagari handwritten character recognition using different feature and classifiers. In: Proceedings of the 10th international conference on document analysis and recognition (ICDAR), pp 1111–1115

Pal U, Roy RK, Kimura F (2010) Bangla and English city name recognition for Indian postal automation. In: Proceedings of the 20th international conference on pattern recognition (ICPR), pp 1985–1988

Park J, Govindaraju V, Srihari SN (2000) OCR in a hierarchical feature space. IEEE Trans Pattern Anal Mach Intell (PAMI) 22(4):400–407

Pasha S, Padma MC (2015) Handwritten Kannada character recognition using wavelet transform and structural features. In: Proceedings of the international conference on emerging research in electronics, computer science and technology, pp 346–351

Patel BC, Kayasth MM (2017) Recognition of offline handwritten Gujarati numerals. I Manag J Inf Technol 6(1):14

Prasad JR (2014) Handwritten character recognition: a review. Int J Comput Sci Netw 3(5):340–351

Prasad SD, Kanduri Y (2016) Telugu handwritten character recognition using adaptive and static zoning methods. In: Proceedings of the IEEE students technology symposium, pp 299–304

Prasad JR, Kulkarni UV, Prasad RS (2009) Offline handwritten character recognition of Gujarati script using pattern matching. In: Proceedings of the 3rd international conference on anti-counterfeiting, security, and identification in communication (ASID), pp 611–615

Prasanth L, Babu JV, Sharma RR, Rao PGV, Manadalapu D (2007) Elastic matching of online handwritten Tamil and Telugu scripts using local features. In: Proceedings of the 9th international conference on document analysis and recognition (ICDAR), vol 2, pp 1028–1032

Purkait P, Chanda B (2010) Off-line recognition of handwritten Bengali numerals using morphological features. In: Proceedings of the 12th international conference on the frontiers of handwriting recognition (ICFHR), pp 363–368

Rahiman MA, Shajan A, Elizabeth A, Divya MK, Kumar GM, Rajasree MS (2010) Isolated handwritten Malayalam character recognition using HLH intensity patterns. In: Proceedings of the 2nd international conference on machine learning and computing (ICMLC), pp 147–151

Raj A (2015) An optical character recognition of machine printed oriya script. In: Proceedings of the 3rd international conference on image information processing, pp 543–547

Rajashekararadhya SV, Ranjan PV (2008a) Neural network based handwritten numeral recognition of Kannada and Telugu scripts. In: Proceedings of the 10th IEEE international conference on TENCON, pp 1–5

Rajashekararadhya SV, Ranjan PV (2008b) Efficient zone based feature extraction algorithm for handwritten numeral recognition of four popular South Indian scripts. J Theor Appl Inf Technol 4(12):1171–1181

Rajashekararadhya SV, Ranjan PV (2009a) Zone based feature extraction algorithm for handwritten numeral recognition of Kannada script. In: Proceedings of the IEEE international conference on advance computing conference (IACC), pp 525–528

Rajashekararadhya SV, Ranjan PV (2009b) Handwritten numeral/mixed numerals recognition of South Indian scripts: the zone based feature extraction method. J Theor Appl Inf Technol 7(1):63–79

Rajput GG, Hangarge M (2007) Recognition of isolated handwritten Kannada numerals based on image fusion method. In: Proceedings of the international conference on PReMI, pp 153–160

Ramakrishnan AG, Shashidhar J (2013) Development of OHWR system for Kannada. VishwaBharat@tdil 39–40:67–95

Rampalli R, Ramakrishnan AG (2011) Fusion of complementary online and offline strategies for recognition of handwritten Kannada characters. J Univers Comput Sci (JUCS) 17(1):81–93

Reddy GS, Sharma P, Prasanna SRM, Mahanta C, Sharma LN (2012a) Combined online and offline assamese handwritten numeral recognizer. In: Proceedings of the 18th national conference on communications (NCC-2012), IIT Kharagpur

Reddy GS, Sarma B, Naik RK, Prasanna SRM, Mahanta C (2012b) Assamese online handwritten digit recognition system using hidden Markov models. In: Proceedings of the workshop on document analysis and recognition (DAR), pp 108–113

Roy K, Pal U (2006) Word-wise hand-written script separation for Indian postal automation. In: Proceedings of the 10th international workshop on frontiers in handwriting recognition (IWFHR), pp 521–526

Roy K, Vajda S, Pal U, Chaudhuri BB (2004) A system towards Indian postal automation. In: Proceedings of the 9th international workshop on frontiers in handwriting recognition (IWFHR), pp 580–585

Roy K, Pal T, Pal U, Kimura F (2005) Oriya handwritten numeral recognition system. In: Proceedings of the 8th international conference on document analysis and recognition (ICDAR), pp 770–774

Sarma B, Mehrotra K, Naik RK, Prasanna SRM, Belhe S, Mahanta C (2013) Handwritten assamese numeral recognizer using HMM and SVM classifiers. In: Proceedings of the 19th national conference on communications (NCC), IIT Delhi

Sastry PN, Krishnan R, Ram BVS (2010) Classification and identification of Telugu handwritten characters extracted from palm leaves using decision tree approach. ARPN J Appl Eng Appl Sci 5(3):22–32

Sastry PN, Lakshmi TRV, Rao NVK, Rajinikanth TV, Wahab A (2014) Telugu handwritten character recognition using zoning features. In: Proceedings of the international conference on IT convergence and security, pp 1–4

Schantz HF (1982) History of OCR, optical character recognition. Recognition Technologies Users Association

Schomaker L (2007) Retrieval of handwritten lines in historical documents. In: Proceedings of the 9th international conference on document analysis and recognition (ICDAR), pp 594–598

Schomaker L, Segers E (1999) Finding features used in the human reading of cursive handwriting. Int J Doc Anal Recognit (IJDAR) 2:13–18

Sethi K, Chatterjee B (1976) Machine recognition of constrained hand-printed Devanagari numerals. J Inst Electr Telecom Eng 22:532–535

Shahin AA (2017) Printed Arabic text recognition using linear and nonlinear regression. Int J Adv Comput Syst Appl 8(1):227–235

Sharma DV, Lehal GS (2006) An iterative algorithm for segmentation of isolated handwritten words in Gurmukhi script. In: Proceedings of the 18th international conference on pattern recognition (ICPR), vol 2, pp 1022–1025

Sharma DV, Lehal GS (2009) Form field frame boundary removal for form processing system in Gurmukhi script. In: Proceedings of the 10th international conference on document analysis and recognition (ICDAR), pp 256–260

Sharma A, Kumar R, Sharma RK (2008) Online handwritten Gurmukhi character recognition using elastic matching. In: Proceedings of the congress on image and signal processing, pp 391–396

Sharma DV, Lehal GS, Mehta S (2009) Shape encoded post processing of Gurmukhi OCR. In: Proceedings of the 10th international conference on document analysis and recognition (ICDAR), pp 788–792

Sharma DV, Jhajj P (2010) Recognition of isolated handwritten characters in Gurmukhi script. Int J Comput Appl 4(8):9–17

Shelke S, Apte A (2011) A multistage handwritten Marathi compound character recognition scheme using neural networks and wavelet features. Int J Signal Process Image Process Pattern Recognit 4(1):81–94

Singh A, Bacchuwar K, Bhasin A (2012) A survey of OCR applications. Int J Mach Learn Comput 2(3):314–318

Sonkusare M, Sahu N (2016) A survey on handwritten character recognition (HCR) techniques for English alphabets. Adv Vis Comput Int J 3(1):1–12

Sopon P, Suksamer T, Polpinij J, Chamchong R (2017) A framework for Thai text retrieval using speech. In: Proceedings of the 6th international conference on computing and informatics, pp 517–522

Sreeraj M, Idicula SM (2010) k-NN based on-line handwritten character recognition system. In: Proceedings of the 1st international conference on integrated intelligent computing (ICIIC), pp 171–176

Srihari SN, Leedham G (2003) A survey of computer methods in forensic handwritten document examination. In: Proceedings of the 11th international graphonomics society conference (IGS), pp 278–281

Sundaram S, Ramakrishnan AG (2008) Two dimensional principal component analysis for online character recognition. In: Proceedings of the 11th international conference on frontiers in handwriting recognition (ICFHR), pp 88–94

Sundaram S, Ramakrishnan AG (2013) Attention-feedback based robust segmentation of online handwritten isolated Tamil words. ACM Trans Asian Lang Inf Process (TALIP) 12(1):4

Sundaram S, Ramakrishnan AG (2014) Performance enhancement of online handwritten Tamil symbol recognition with reevaluation techniques. Pattern Anal Appl (PAA) 17(3):587–609

Sunija AP, Rajisha RM, Riyas KS (2016) Comparative study of different classifiers for Malayalam dialect recognition system. In: Proceedings of the international conference on emerging trends in engineering, science and technology, pp 1080–1088

Swaileh W, Lerouge J, Paquet T (2016) A unified French/English syllabic model for handwriting recognition. In: Proceedings of the 15th international conference on frontiers in handwriting recognition, pp 536–541

Tran DC, Franco P, Ogier J (2010) Accented handwritten character recognition using SVM-application to French. In: Proceedings of the 12th international conference on frontiers in handwriting recognition (ICFHR), pp 65–71

Tripathy N, Pal U (2004) Handwriting segmentation of unconstrained Oriya text. In: Proceedings of the 9th international workshop frontiers in handwriting recognition (IWFHR), pp 306–311

Tsai C (2016) Recognizing handwritten Japanese characters using deep convolutional neural networks, Report, pp 1–7

Venkatesh N, Ramakrishnan AG (2011) Choice of classifiers in hierarchical recognition of online handwritten Kannada and Tamil aksharas. J Univers Comput Sci (JUCS) 17:94–106

Wang X, Govindaraju V, Srihari S (2000) Holistic recognition of handwritten character pairs. Pattern Recognit 33(12–33):1967–1973

Yap PT, Paramesran R (2003) Image analysis by Krawtcouk moments. IEEE Trans Image Process 12(11):1367–1377

Zhang XZ, Yan CD, Liu XY (1990) Feature point method of Chinese character recognition and its application. J Comput Sci 5(4):305–311

Zhu B, Zhou XD, Liu CL, Nakagawa M (2010) A robust model for on-line handwritten Japanese text recognition. Int J Doc Anal Recognit (IJDAR) 13(2):121–131