



A comprehensive survey on word recognition for non-Indic and Indic scripts

Harmandeep Kaur¹ · Munish Kumar¹

Received: 25 September 2017 / Accepted: 18 July 2018 / Published online: 25 July 2018
© Springer-Verlag London Ltd., part of Springer Nature 2018

Abstract

The term handwriting recognition is used to describe the capability of a computer system to transform human handwriting into machine processable text. Handwriting recognition has many applications in various fields such as bank-cheque processing, postal-address interpretation, document archiving, mail sorting and form processing in administration, insurance offices. A collection of different scripts is employed in writing languages throughout the world. Many researchers have done work for handwriting recognition of various non-Indic and Indic scripts from the most recent couple of years. But, only a limited number of systems are offered for word recognition for these scripts. This paper presents an extensive systematic survey of word recognition techniques. This survey of word recognition is classified broadly based on different scripts in which a word is written. Experimental evaluation of word recognition tools/techniques is presented in this paper. Different databases have been surveyed to evaluate the performance of techniques used to recognize words, and the achieved recognition accuracies have been reported. The efforts in two directions (non-Indic and Indic scripts) are reflected in this paper. We increased awareness of the potential benefits of word recognition techniques and identify the need to develop an efficient word recognition technique. Recommendations are also provided for future research. It is also observed that the research in this area is quietly thin and still more research is to be done, particularly in the case of word recognition of printed/handwritten documents in Indic scripts.

Keywords Word Recognition · Holistic approach · Non-Indic scripts · Indic scripts

1 Introduction

The field of pattern recognition contributed up to a great extent in the machine vision applications. Handwriting recognition is a part of the area under pattern recognition community. Handwriting recognition is the technique by which the computer system can recognize characters and other symbols written by individuals using natural handwriting. All this should be made to keep the records in the computer system in digital form [64] that can be used for future references. So, basically, to interact with a computer system or to exchange information with the computer, users have to input the required data into the system. The process of converting

textual symbols on a paper to a machine process-able format is known as optical character recognition (OCR) which is the core of the field of document analysis system (DAS). Thus, it plays an important role in transformation of paper based society to paperless electronic information society. OCR technology for Indian documents is in emerging stage and most of these Indian OCR systems can read the documents written in only a single script. As per the tri-scripts principle of Indian constitution, every state Government has to produce an official document containing a national script (Devanagari), official script (Roman) and the state script (or regional script). For example, an official document of Punjab state contains Devanagari, Roman and Gurmukhi scripts. The processing of such complex multi-script documents is a challenging problem for OCR researchers. Since a word is essentially a sequence of characters, a natural approach to word recognition is to segment the word into characters and recognize the individual characters using OCR systems. In most of the applications, it is reasonable to suppose that a lexicon is provided. The lexicon can be either static or

✉ Munish Kumar
munishcse@gmail.com

Harmandeep Kaur
harmandeepk08@gmail.com

¹ Department of Computational Sciences, Maharaja Ranjit Singh Punjab Technical University, Bathinda, Punjab, India

generated dynamically. The word recognition model says that words are recognized as complete units, is the oldest model in the psychological literature. The general idea is the fact that we see words as a complete pattern rather than the sum of letter parts. Some claim that the information employed to recognize a word is the pattern of ascending, descending, and neutral characters. Another formulation is to use the envelope created by the outline of the word as depicted in Fig. 1a, b. Word patterns are recognizable to us as an image because we have seen each of the patterns many times. Cattell [8] was the first psychologist to propose this as a model of word recognition. He is recognized as a great founder of the field of psycholinguistics, which includes the scientific study of reading.

He had discovered a fascinating effect that is known as “Word Superiority Effect.” He presented the letter and word stimuli to subjects for a very brief period of time (5–10 ms) and found that subjects were more accurate at recognizing the words than the letters. He concluded that subjects were more accurate at recognizing words in a short period of time because whole words are the units that we recognize. The second key piece of experimental data to support the word shape model is that lowercase text is read faster than uppercase text. Word Recognition is the ability of a recognition system to recognize paper-based printed/handwritten words correctly and effortlessly. Word recognition is more important than character recognition because it is helpful for rapid and fluent reading. It will also save data electronically; thus, files will be saved digitally in an efficient way, and it reduces the number of character errors associated with OCR engine. Woodworth [78] was the first to report this finding in his influential textbook “*Experimental Psychology*.” This finding has been confirmed more recently by Smith [66] and Fisher [19]. Participants were asked to read comparable passages of text, half completely in uppercase text and half presented in the standard lowercase text. In each study, participants read reliably faster with the lowercase text by a 5–10% speed difference. This supports the word shape

model because lowercase text enables unique patterns of ascending, descending, and neutral characters. The shortest lived model of word recognition is that words are read letter-by-letter serially from left to right. Gough [17] proposed this model because it was easier to understand, and far more testable than the word shape model of reading. This model is consistent with Sperling’s [67] finding that letters can be recognized at a rate of 10–20 ms per letter. Bouwhuis and Bouma [6] developed a model of word recognition based on the probability of recognizing each of the letters within a word. They conclude that “word shape might be satisfactorily described in terms of the letters in their positions.”

Mori et al. [43] have reviewed research and development of OCR (Optical Character Recognition) systems from a historical point of view. The research and development part of OCR systems is divided into two approaches, namely template matching and structure analysis. In general, it is difficult to separate these two approaches and thus can be merged into one wide stream. On the other hand, the commercial OCRs can be categorized into three generations. The first generation can be characterized by the constrained letter shapes which the OCRs read. The second generation can be characterized by machine-printed and hand-printed character recognition capabilities. The third generation can be characterized by the recognition capabilities of poor print quality characters and hand-printed characters for a large category character set, such as Chinese characters. Some techniques applied to OCR such as expert systems, neural networks and some open problems are also discussed in this paper. Steinherz et al. [68] have presented a survey on off-line cursive word recognition field. In this survey, various methods for the word recognition system have been discussed in view of two most important properties of the recognition system, namely the size and nature of the lexicon involved, and whether or not a segmentation stage is present. The field of off-line cursive word recognition is classified into three categories, namely segmentation-free, segmentation-based and perception-oriented approaches. In segmentation-free method, without performing segmentation, the best interpretation of an observation sequence derived from the word image can be found. For finding the most likely interpretation, minimum edit-distance algorithm is used, that is implemented by classic dynamic programming tools and unique hidden Markov models (HMMs). Segmentation-based method is based on finding the best match between blocks of primitive segments and a word’s letters. In order to find the optimal solution, either dynamic programming, graphical models or HMMs can be used. Perception-oriented approach performs a human-like reading technique, in which anchor features all over the word are used to bootstrap a few candidates for a final evaluation phase.

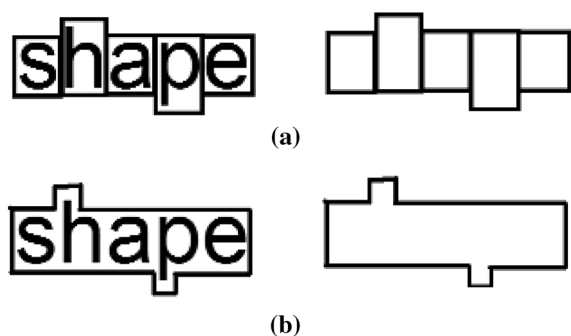


Fig. 1 **a** Word shape recognition, **b** Word shape recognition using the envelope around the word

Vinciarelli [75] has presented a survey on off-line cursive script word recognition by considering two parts. The first part deals with the general aspects of cursive script recognition (CSR), and the second part illustrates the main applications of CSR. The mainly used recognition approaches are the explicit segmentation followed by dynamic programming and the implicit segmentation followed by Hidden Markov Model (HMM). The application areas of CSR include bank check and postal-address reading. It has also been shown that the human reading inspired systems are effective only in applications containing a lexicon of 25–30 words. Obaidullah et al. [46] have presented multi-level script identification at page, block, line and word level in the same document. Two features, namely script dependent and script independent features, are considered, and they have used two classifiers, namely multilayer perceptron (MLP) and random forest (RF) to investigate the script recognition performance considering all the levels. In the dataset, eleven Indic scripts are considered, namely Bangla, Devanagari, Gujarati, Gurmukhi, Kannada, Malayalam, Oriya, Roman, Tamil, Telugu and Urdu. The dataset includes 440 pages, 2200 blocks, 3300 lines and 6600 words where the 5 blocks, 7.5 lines and 15 words are generated from a single page. The results reveal that MLP performs best in all the situations and the line-level data gives most consistent results, followed by page, block and word-level. Regarding the feature set, script independent features are suitable at word-level and script-dependent features are suitable when document size is considerably larger. Kumar et al. [37] have presented a survey based on isolated character/numeral recognition of non-Indic and Indic scripts. In this survey, they have also examined major challenges/issues of character/numeral recognition. Lehal and Singh [39] presented a system for recognition of machine-printed Gurmukhi text which was the initial step for Gurmukhi text recognition. Their recognition system works at the sub-character level as shown in Fig. 2. Handwritten word can consist of cursive characters, discrete characters or a mixture of both as depicted in Fig. 3.

This paper presents a systematic survey to analyze and report the findings in word recognition. Systematic surveys are the time consuming process, but provide transparent and

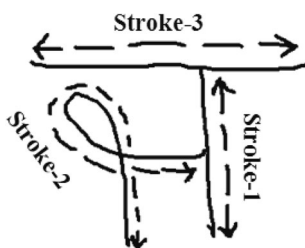


Fig. 2 Handwritten character “Jajja” of Gurmukhi script with three strokes

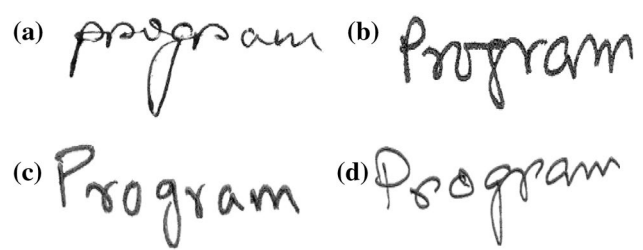


Fig. 3 A sample of Roman word. a Cursive, b touching discrete, c purely discrete, d mixture of cursive and discrete words

comprehensive view of ongoing research, and can be used to identify a number of research avenues. This survey identifies different key areas of research on word recognition, discusses the concepts, research methods used and major findings.

2 Motivations

Indian enterprises, such as public sectors like, banking, financial sector, insurance, are unable to provide their services to a majority of the country’s population. There are many reasons for this, such as low literacy rate of the population (65%), issues with accessibility, usability and availability of these services. Even with the advent of Information Technology (IT) and success of Mobile Technology (MT), many services have not been able to reach the common people. Many organizations in India still use paper (e.g., forms) as part of their workflow. This heavy dependence on paper necessitates the manual processing resulting in poor and slow services. Though technologies like automated forms-processing, optical character recognition and handwritten word recognition in Roman script were successfully applied in other countries, they have not created a significant impact in public offices in India as they can only address a small Roman literature portion. According to the national readership survey 2002, in India less than 10% of the newspapers are read in English as compared to 33% that are read in Hindi (out of 150 million readers) and the rest are read in other local languages. There are many local languages and written scripts in the India (22 languages and 12 scripts). Hence, there is a clear need for commercial off-line handwriting recognition engines in local scripts such as Bengali, Devanagari, Gurmukhi, Tamil, Telugu for providing services to a local-language-literate population. This could empower Indian enterprises to start deploying forms-processing technologies with handwriting in local scripts. It will also be possible to extend services to remote locations in India via mobile phones with cameras, where filled application forms can be submitted by just clicking and sending the images of papers. Moreover, handwritten word

recognition can be used in many applications like postal-address identification, writer's handwriting identification, bank-checkue recognition, signature verification in banks, historical documents, identifying the words in inscription. It is not possible to save the historical documents, writer's books for many years in the original format. But, once it is digitized, then it's very easy to use such documents for the generation to generations. Because of the improvement in the technology of the past few decades, the older historical documents are stored in the digitized form. Hence, it can be helpful for the future generation for extracting, modifying and storing the data. Off-line handwritten word recognition is the best approach to achieve this goal. So, handwritten word recognition has become a potential leading research area in the field of document analysis and recognition.

3 Background

In this section, we define the word recognition, types of handwritten word recognition, the different approaches of handwritten word recognition and the factors leading to word recognition. After that, we presented the advantages and disadvantages of word recognition.

3.1 Word recognition

Word recognition, according to Literacy Information and Communication System (LINCS) is “the ability of a reader to recognize written words correctly and virtually effortlessly.” The article “The Science of Word Recognition” says that “evidence from the last 20 years of work in cognitive psychology suggests that we use the letters within a word to recognize a word.” Other theories have been put forth proposing the mechanisms by which words are recognized in isolation. Word recognition is a manner of reading based upon the immediate perception of what word a familiar grouping of letters represents. Handwriting recognition is the ability of a computer to receive and interpret intelligible handwritten input from sources such as paper documents, photographs, touch-screens and other devices. Handwriting recognition principally entails OCR systems. However, complete handwriting recognition system also handles formatting, performs correct segmentation into characters and find the most possible words.

3.2 Types of word recognition

Handwriting word recognition (HWR) is divided into two parts, namely off-line word recognition and online word recognition.

- **Off-line handwritten word recognition:** Off-line handwritten word recognition deals with the recognition of handwritten words after it was written. It involves the processing of documents containing scanned images of a text written by a user, generally on a sheet of paper. In this kind of system, words are digitized to obtain two dimensional images. Stroke information is not available in the off-line handwriting recognition.
- **Online handwritten word recognition:** In online handwritten word recognition the trajectories of pen tip movements are recorded and analyzed to identify intended information. With the latest technological advancements, it is very common to write on ordinary paper and immediate transmission of handwritten annotations to a remote server is possible. In online handwritten word recognition, the writing is done with a special pen on an electronic notepad or a tablet and where temporal information, such as the position and velocity of the pen along its trajectory, is available to the recognition algorithm. Since most algorithms for online handwritten word recognition attempt to recognize the writing as it is being written, sometimes it is also referred to as “real-time” handwritten word recognition.

3.3 Approaches to word recognition

Handwritten word recognition is an active research area in various fields of machine vision applications. Generally, two main approaches are available in the literature to recognize handwritten words which are listed as below:

- **Segmentation-based approach:** In this approach, initially the words are segmented into characters or pseudo-characters and then the character or pseudo-character model is used for recognition. The success of the recognition module depends much upon segmentation performance. But segmentation itself is often ambiguous and very prone to failure.
- **Segmentation-free approach:** This approach is also known as the holistic approach. The approach treats the word itself as a single entity and it goes for recognition without doing explicit segmentation.

A general framework for word recognition is depicted in Fig. 4.

3.4 Why word recognition

We list some of the reasons behind using the word recognition technique:

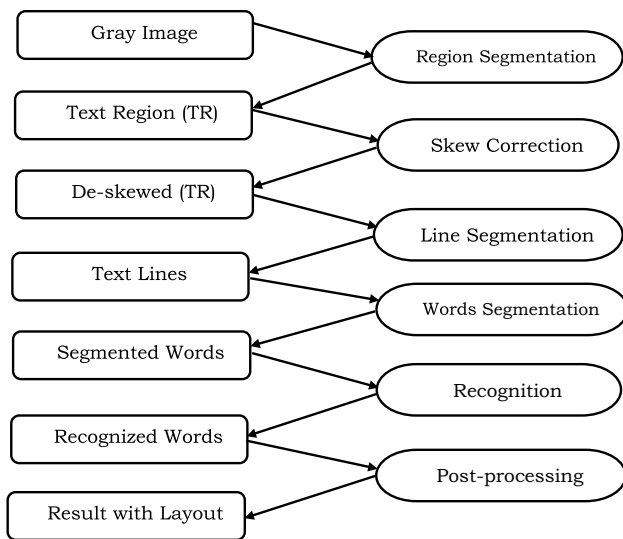


Fig. 4 A general framework for word recognition

- Word recognition system recognizes entire handwritten words or phrases instead of character-by-character, like its predecessor, Optical Character Recognition (OCR).
- Word recognition technology matches handwritten or printed words to a user defined dictionary, significantly reducing character errors encountered in typical character-based recognition engines.
- Touching with nearby letters is a serious problem in the segmentation of handwritten character recognition. This problem can be solved by using a holistic approach to handwritten word recognition as there is no segmentation in this approach.

3.5 Advantages of word recognition

- Word recognition eliminates a large proportion of the manual data entry of handwritten documents that, in the past, could only be keyed by a human, creating an automated workflow.
- It saves data electronically; thus, files will be saved digitally in an efficient way.
- It reduces the number of character errors associated with OCR engine.
- The segmentation-free approach to word recognition is achieving the best performance on standard benchmarks.
- Once the historical document is digitized, then it's very easy to use such documents for the generation to generations.
- WritePad handwriting recognition system is developed for iPhone, iPod and iPad Touch devices that can recognize all styles of writing.

3.6 Disadvantages of word recognition

- Unique styles of writing make it a difficult process to recognize a word.
- For methods based on a matching of a prototype, a new prototype must be created with each added word.
- Due to different types of languages, the processing of complex multi-script documents is a challenging problem for OCR researchers.
- Degraded images of text, spacing of letters or words makes it an additional problem to word recognition.

4 Issues and challenges

Off-line handwritten word recognition is one of the most difficult tasks compared to online word recognition approach. This is because of the various writing styles of individuals, thickness of the pen, environment depend on the situation of the writer, etc. Though so many research articles are available in the literature, still handwritten word recognition is an open problem, especially in the domain of feature extraction and classification methodology. Large number of character set present in the particular language makes it an open problem for the researchers. In handwritten documents, it is often not easy to recognize punctuation marks, as they are not precisely rendered during writing. Thus, a comma may be mistaken for a character owing to both its size and location. In handwritten words, it would often be very difficult to distinguish the confusing words like “clean” from “dean,” when wrote in a cursive mode. Contextual interpretation would be required to resolve this type of confusion. Again, due to the diverse writing styles of people, sometimes unusual ligatures connecting adjacent characters often add confusion to the identity of the word. This is especially problematic with words that have “w,” “u,” “v,” etc. The artifacts of the complex interactions between hardware equipment and subsequent operations are scanning and binarization for off-line handwritten word recognition.

5 Survey protocol

A systematic survey of word recognition has been reported in this section. The steps included in the survey include: development of a survey protocol, conducting the survey, analyzing the results, reporting the results and discussion of findings.

5.1 Planning the survey

The survey protocol includes the research questions framework, the databases searched, methods used to identify and

assess the evidence. Conducting the survey comprises identification of primary studies, applying inclusion and exclusion criteria and producing the results. Electronic databases were extensively searched, and its studies are reported.

5.2 Research questions

The main goal of this systematic survey was to determine and classify the existing literature focusing on word recognition techniques. To plan the survey, a set of research questions was needed. Table 1 depicts the specific research questions and sub-questions.

5.3 Sources of information

A broad perspective is necessary for an extensive and broad coverage of the literature. Before starting the research, an appropriate set of databases must be chosen to increase the probability of finding highly relevant articles. Following databases were searched:

- CMATERdb1.2.1
- CENPARMI
- CEDAR (Centre of Excellence in Document Analysis and Recognition) benchmark database
- ICDAR Robust Reading Competition datasets (ICDAR 2003, ICDAR 2011 and ICDAR 2013)
- Google Street View Text dataset (SVT)
- FARSA
- IFN-ENIT
- Ibn Sina
- Arabic handwriting dataset AHDB
- MCW database
- Online handwritten Mongolian word database named MRG-OHMW
- IRONOFF, SRTP-Cheque and AWS

5.4 Inclusion and exclusion criteria

In the initial phase, irrelevant papers were excluded manually based on titles. Studies were able to claim for inclusion in the survey if their focus of study was word recognition. Studies of both research scholars and professional software developers were included. The systematic survey included qualitative and quantitative research studies, published up to and including 2017 starting from the initial date of the digital library to make the database search comprehensive. Only studies written in English were included. We included technical reports in our study. Studies were excluded if their focus was not on word recognition. Research papers repeating in different e-resources and databases were individually excluded to ensure our research database remained normalized.

5.5 Quality assessment

After using the inclusion and exclusion criterion to select relevant papers, a quality assessment was performed on the remaining papers. Since the field is eclectic, a large number of different journals and conferences include research papers of our interest. Using the quality assessment as per “Appendix 1,” all of the included papers contain the high-quality word recognition research, providing additional confidence in the database of selected papers. In the quality assessment form (Appendix 1), the higher level question in Sect. 1 sets the basis for screening the study. After the research paper was included, the paper was studied for classification based on questions in Sect. 2. Then we proceeded to Sects. 3 and 4.

5.6 Data extraction

Appendix 2 identifies the guidelines for data extraction from all the 82 studies included in this systematic literature

Table 1 Research questions, sub-question and motivation

Research question	Motivation
(1) What is the current status of word recognition? (1.1) What methods of word recognition are used? (1.2) What tools are available for word’s feature detection, what methods do they use, what type of scripts do they address? (1.3) Which studies have evaluated methods/tools and with what results?	It helps in understanding word detection techniques. Various techniques used in word recognition are reported. Various tools/techniques for word detection developed till date are mentioned with their usage. The research question explores the studies which evaluated/compared different word recognition techniques. The number of studies for each type of script is also reported
(2) Key sub-areas (2.1) What are the important areas related to word recognition, the number of studies in each classified area and their findings? (2.2) A time based count to show how the area has evolved over time	It helps in knowing the type of study carried out in the article. It is important to understand the number of studies for each sub-area which helps in identifying key areas for further research. Time based count shows how the key area has evolved over time
(3) What is the subject system used? (3.1) What is the size of the database used in recognizing words? (3.2) Whether the system is open source or commercial?	It will help in creating the database on which the word recognition research can be carried out. It is a move towards benchmarking and standardization of comparative analysis studies

survey. The data extraction form was designed when we started the information gathering process which is sufficient to address the research questions framed. Quality assessment form in “Appendix 1” sets the basis for inclusion/exclusion criteria of the study. To some extent, our quality assessment form and data extraction form overlapped. When we started the systematic survey, we experienced many problems. It was difficult to extract all the relevant data (Appendix 2) from many studies. Due to this, it was necessary to contact many researchers to find the required details which we were not able to infer from the research paper.

The data extraction procedure can be summarized as:

- Initially, the authors surveyed all of the papers and extracted data from all the primary studies.
- To check the consistency of data extraction, another researcher performed data extraction on a random sample of the primary studies and the results were cross-checked.
- If there were any disagreement when papers were cross-checked, consensus meetings among the authors were used to resolve them.

6 Reported work

After including all the relevant papers related to word recognition, a detailed analysis of these papers has been done on the basis of work reported for the non-Indic and Indic scripts.

6.1 Non-Indic scripts

Script wise work related to word recognition of non-Indic scripts is summarized in this section: Dasgupta et al. [11] presented a holistic approach for off-line handwritten cursive word recognition using directional features based on Arnold transformation. Extraction of diagonal features depends on the stroke orientation distribution of cursive word. They partitioned the image in a fashion similar to a quad-tree. That means, first, the image is partitioned horizontally and vertically into four equal non-overlapping blocks. Then each block or sub-image is further partitioned into four non-overlapping blocks. This partitioning continues up to a desired level. Now each sub-image in each level of partitioning is scrambled by Arnold transformation through one period, and distribution of stroke orientation of the Arnold transformed image is computed by the Hough transform. However, to the best of their knowledge Arnold transformation has not been used for directional feature extraction from any type of script data before. Besides this feature, some other directional shape features are also used to form a feature vector. Finally, words are recognized using a well-known

multi-class support vector machine (SVM) classifier. The proposed method is tested on a benchmark handwritten word database CENPARMI database of legal amounts written in English. The database contains English legal words collected from 2500 handwritten Canadian bank cheques written by about 800 persons including the staff of Concordia University. The data set contains 32 different words or classes, including “hundred,” “thousand,” “and,” “only” and “dollar.” They have partitioned it into training (5230 words) and test (2514 words) datasets and an overall accuracy of 87.19% is achieved.

Singh et al. [65] proposed a technique for the recognition of legal amount words written in English Script based on template matching technique using correlation coefficient. The legal amount is the value of a cheque written in words. This is a holistic approach, i.e., a word is treated as a single entity or an object, and not as a sequence of letters. They rather generated template of each word under consideration as its prototype and compared the test word with these prototypes to recognize the word. Once the handwritten word recognition system receives a test word image, then a cover image of the test word is generated using the process. Then similarity of this cover image with templates is measured as the distance between them. The word corresponding to template for which distance is minimum is said to be recognized word provided this minimum distance is less than a predefined threshold. As there is no reliable online database available for this application so they have developed the database of 61 words, the combination of which can represent any legal amount written in words in Indian bank cheque. In this experiment, they have used handwritten word samples of 63 writer’s samples where each writer writes each of the words five times resulting in $(61 \times 63 \times 5 = 19,215)$ 19,215 word samples and word recognition accuracy of 76.4% has been achieved. Zhang et al. [79] have presented new algorithms and models for object level video advertising. For comparison, they employed the GA, a stochastic global optimization algorithm, to optimize the model with an appropriate encoding scheme for chromosomes and using global information. Zhang et al. [80] presents a character-level sequence-to-sequence learning method. In their work, they embedded Recurrent Neural Network (RNN) into an encoder-decoder framework and generate character-level sequence representation as input. This method reads quantized characters into the translation system, instead of using a predefined vocabulary with a limited number of words. Oyedotun and Khashman [47] applied deep learning to the problem of hand gesture recognition for the whole 24 hand gestures obtained from the Thomas Moeslund’s gesture recognition database. They trained convolutional neural networks (CNNs) and stacked de-noising auto-encoders (SDAEs) on a public database; and recognition rates of 91.33 and 92.83% are obtained, respectively, using test data that are not part of the training data.

Caesar et al. [7] presented a system for the recognition of images of handwritten cursive words. All features of the described system are based on a symbolic representation of the contour and the skeleton. They used the Hidden Markov Models (HMM) to handle the enormous variance of handwriting. Before applying the classifier, they defined three steps. In a first step, the vector quantizer is adapted to the statistics of feature vectors. In the second step, the linear discriminant analysis is adapted to clusters directly defined by hidden Markov models. After that the system is ready to adjust the HMM parameters by a semi-supervised learning. The parameters of the model emission and transition probabilities are estimated from the sample set by the forward/backward adaptation technique. The HMM recognizer was trained by a learning set of US and German city names of approximately 15,000 word images during vector quantization and approximately 10,000 images during fine adaptation of the second pass of the training. The recognition results are 91.0% with a test set of approximately 1280 US city names and 85% with a test set of approximately 200 German city names. Fusion is one of the powerful methods for improving recognition rates produced by various techniques. Verma et al. [73] proposed a fusion of three handwritten word recognitions using a Modified Borda Count (MBC). In MBC they have added three new components, i.e., rank, confidence value of each word in the lexicon based on character confidences, compatibility scores and the third component is weighted variable in which higher weight is assigned to the techniques with higher recognition rates and a lower weight is assigned to the techniques with lower recognition rates. Three handwritten word recognition techniques used are: MUMLP system based on over-segmentation, a multilayer perceptron trained using the back-propagation algorithm and dynamic programming. GUMLP system based on the heuristic segmentation algorithm used to locate perspective segmentation points in handwritten words. MURBF system is based on radial basis function neural network. The experiments were conducted on cursive handwritten words taken from the CEDAR (Centre of Excellence in Document Analysis and Recognition) benchmark database. The database contains real world zip codes, city and state names from handwritten postal envelopes which were obtained from United States Postal Services (USPS). The database is divided into training and testing sets of 3106 and 317 words, respectively. The proposed Borda count achieved the highest recognition rate 91.0%, which is much better than any individual technique.

Su and Lu [69] proposed a novel text recognition technique that performs word level recognition without character segmentation. The overall system consists of three components. In the first component, a word image is converted into a sequential feature, i.e., sequences of column feature based on HOG (Histograms of Oriented Gradients) features

with different parameter settings. In total, they extracted two feature sets for training. In the second component, two multilayer recurrent neural network (RNN) model with bidirectional long short-term memory (LSTM) and connectionist temporal classification (CTC) is trained to classify the two sets of sequential data. Finally, an ensemble technique is used that combines the outputs of multiple RNNs to produce improved word recognition accuracy. The proposed method has been tested on four datasets, including three ICDAR Robust Reading Competition datasets (ICDAR 2003, ICDAR 2011 and ICDAR 2013) that consist of scene images captured in different environments, and Google Street View Text dataset (SVT) that mainly consists of images of the signboards and shops' names in outdoor environments. Another three datasets are also included for training: ICDAR Born Digital Image Dataset (BDI), Sign Recognition Dataset (SRD) and IIIT5K Dataset. They evaluate recognition accuracy on testing data with a lexicon created from all the words in the test set [as denoted by ICDAR03 (FULL) and ICDAR11 (FULL)], as well as with lexicon consisting of 50 random words from the test set [as denoted by ICDAR03 (50) and ICDAR11 (50)]. For the SVT dataset, they directly adopt the 50-word lexicon. With ICDAR03 (FULL), ICDAR11 (FULL), ICDAR13 (FULL) and SVT data sets the achieved recognition accuracies are 89.0, 87.0, 90.0 and 89.0%, respectively. The accuracies of the proposed technique on the ICDAR-03, ICDAR-11 and SVT datasets drop to 72.0, 69.0 and 70.0%, respectively, when no lexicon is used.

6.1.1 Work on single script

6.1.1.1 Arabic The Arabic script is the writing system used for writing Arabic and several other languages of Asia and Africa. This script is written from right to left in a cursive style and includes 28 letters. It is the second most widely used writing system in the world. Hafiz and Bhat [20] proposed a two-tier hybrid classification scheme to boost the recognition capability of HMM (Hidden Markov Model)-based Arabic Optical Character Recognition (OCR) systems. This is the first attempt to use a hybrid HMM-KNN classifier for Arabic OCR. The approach is segmentation based. The first tier of the hybrid system is a Part-of-Arabic-Word (PAW)-based HMM classifier, which generates the corresponding log-probabilities for each PAW image. The emitted PAW vector is converted into an integer vector, by assigning to each scalar, a serial number after a simple serial-wise look-up in the overall PAW list for the dataset. This process is performed for both training set as well as the testing set. The second tier is a k -nearest neighbor (k -NN) classifier which receives its inputs from the first tier and assigns classes to the emitted PAWs from tier one. The database has been obtained from the IFN-ENIT database

of Arabic words that comprise of three sets of randomly selected images. Set A had 600 word images (15 words with a total of 39 PAWs). Set B had 400 word images (10 words with a total of 32 PAWs). Set C had 200 word images (5 words with a total of 19 PAWs). Thirty images were used per word for training and ten images were used per word for testing. The proposed hybrid scheme achieved the classification accuracy of 82.67, 86 and 94% on sets A, B and C, respectively. To enhance the performance of IKRAA, a system for recognition of handwritten Arabic words based on a transparent neural network (TNN-DF), Cheikh and Kacem [9] investigated the use of three NN-MLPs (neural network-multilayer perceptrons) at the training stage. TNN-DF uses structural features for the global description and the recognition of the words and Fourier descriptors (DF) for the local description of some letters only if an ambiguity exists. Their approach consists, firstly, in the conception, then the training of three mono-layer perceptrons using the “gradient retro-propagation” rule and secondly, in the migration of the weights obtained from the network training toward the TNN-DF. The three proposed networks are dedicated to the training of letters, parts of words and words from the descriptions of the city names. This description is based on global morphological characteristics. The first network, called NN-MLP1, has 50 neurons in the input layer and 113 neurons in the output layer (as many neurons as Arabic letters). The second network, called NN-MLP2, has 113 neurons in the input layer (as many neurons as letters) and 164 neurons in the output layer (as many neurons as PAWs). The third network, called NN-MLP3, has 164 neurons in the input layer (as many neurons as PAWs) and 50 neurons in the output layer (as many neurons as Tunisian city names). For the development of the TNN-DF, they have used some samples of Tunisian city names extracted from the IFN/ENIT database (15 samples for each of the 50 city names), 500 samples used in the training set and the rest for the recognition. After including the training stage, the TNN-DF offers a recognition score equal to 77.6%.

Khemiri et al. [27] proposed a system for the off-line recognition of handwritten Arabic words based on a Bayesian approach. The proposed system contains three main steps: baseline estimation, feature extraction and word classification. For baseline estimation, they proposed two methods. The first one compares for each word the obtained baseline according to a manually added baseline. Comparisons are done point by point. In the second method, they refer to both manual baseline and IFN-ENIT’s ground-truth baseline and comparisons are made segment by segment. Once the baseline estimated, it makes possible to extract the effective word central band. Different structural features such as ascenders, descenders, loops and diacritic, are extracted from word’s image, taking into account the morphology of handwritten Arabic words. For word modeling and recognition,

the extracted features are used as input to some variants of Bayesian networks, namely, Naive Bayes (NB), tree augmented Naive Bayes network (TAN), horizontal and vertical hidden Markov model (VH-HMM) and dynamic Bayesian network (DBN). Experiments are carried on the IFN/ENIT database. They considered 83 classes, 7881 word samples (2/3 for the training and 1/3 for the testing) and the highest rate of 90.02% was achieved by VH-HMM. Septi and Bedda [59] described an automatic recognition system for the handwritten Arabic words. The major problem in the automatic recognition of the cursive Arabic writing is the segmentation in their different constituent. First of all, they classify the city names along with the number of the related components; then they segment every component in characters. For the segmentation of characters, they proposed a method that essentially rests on the skeletons of the images of the cities and the detection of points of branching or crossing (essential points). After the segmentation, they tried to describe the characters by their applicable features (topological features). A dataset of 48 cities of Algeria in shape of Arabic manuscript is considered. For each class, they define a network of neurons of type multilayered perceptrons. The rates of recognition are between 90.00 and 98.33%.

Bouaziz et al. [5] presented an Arabic handwritten word recognition system for wide vocabulary using the analytical approach. After binarization, word images pass through the pretreatment unit for cleaning. Diacritic mark elimination was performed subsequently after safeguarding their coordinates in order to simplify the treatment. Diacritic mark detection is made up of two steps. In the first step, they performed the filtering of the connected components using fixed thresholds to eliminate diacritics, without affecting the connected components corresponding to the shape of the character. The second filter takes into account the location of diacritics relative to the baseband in order to complete detection of the residual diacritics. After these pretreatments, the word images are segmented using a segmentation algorithm based on the detection of ligatures between the characters by analyzing vertical projection profiles. The feature vector contains 128 elements combining structural features families extracted from the original word image to recognize. They adopted the nonlinear support vector machine (SVM) classifier whose kernel is the radial basis function, since it has a learning phase and it produces the classes listed with associated weights as a result. For the treatment of multi-class data, they used the approach “one against one” given its implementation simplicity. After injecting diacritics, the proposed system offers assumptions letters for each recognized character according to the number and the position of diacritics. Thus, the proposed system generates assumptions of words by concatenation of letters assumption. To validate the generated word hypothesis, they used a language dictionary consisting of 6446 words which is helpful to filter the

word hypothesis and keep only valid words. They built a dataset of 500 words written by four writers and achieved the recognition rate of 96.82%.

Khaissidi et al. [26] presented an unsupervised segmentation-free method for spotting and searching query, especially, for images of documents in handwritten Arabic. Histogram of oriented gradient (HOG) is used to detect and extract features of Arabic handwritten documents. Implementation of HOG descriptors is achieved by dividing the image into small connected regions, called cells, and for each cell computing a histogram of gradient directions. Histograms are also normalized based on their energy. The combination of these histograms represents the descriptor. They compressed the descriptors with the product quantization method which provides better performance in a time of descriptor computation. This method reduces the amount of memory needed to store the descriptors and reduce the computational cost of searching for the descriptor. Finally, a better representation of the query is obtained by using the support vector machines (SVM) with a linear function. This method has been evaluated on a large amount of handwritten Ibn Sina datasets and achieved mean average precision of 68.4%. Jayech et al. [23] proposed an approach to the recognition of off-line handwritten Arabic city names based on the dynamic hierarchical Bayesian network (DHBN). The proposed system is the first attempt to experiment with the IFN/ENIT database with the dynamic Bayesian network (DBN). In this system, a non-uniform segmentation approach based on the vertical histogram projection using various width values to put down the segmentation error is proposed to segment the word into characters. After that, they segment the character into frames and cells using a uniform segmentation. The number of frames is fixed empirically, yielding the highest recognition rate, to 3 and the number of cells to 2. After that, they extract the characteristics of each cell using the Zernike and HU moments, which are invariant to rotation, translation and scaling. They used discrete DBNs to process to the next step of preprocessing, which consist in quantizing each continuous feature vector representing a cell to a discrete symbol. Then, the sub-character is estimated at the lowest level of the Bayesian Network (BN), and the character is estimated at the highest level of the BN. Overall Arabic words are processed by a dynamic BN. The developed system has been tested using the IFN/ENIT database which consists of 946 handwritten Tunisian city names and their corresponding postcodes and achieved 82.0% accuracy. Thus, the results show a significant improvement in the recognition rate.

Karim and Kadhm [24, 25] proposed a new framework for a handwriting word recognition system based on neural nets (NN) classifier. The proposed work depends on the handwriting word level, and it does not need character segmentation stage. In the proposed architecture, structural features

consist of zigzag, dots, loops, end points, intersection points and strokes in many directions. Two types of statistical features have been used, namely, connected components feature and zoning features. Two effective transform methods are used for extracting the text features, namely, discrete cosine transform features (DCT) and histogram of oriented gradient (HOG). Feature extraction technique is implemented and gives 405 feature points for each image accordingly. After the feature extraction, the major task is to make the decision to classify the word to which class it belongs and for this purpose Neural Network (NN) classifier is used. An Arabic handwriting dataset has been used for training and testing the proposed system that has 2913 handwriting word images. Each word has 105 images written in a different style. In the handwriting word recognition system, 70% of the dataset is used for training purpose and 30% for testing and it achieved 95.0% recognition accuracy. Karim and Kadhm [24, 25] also proposed a new framework for a handwriting word recognition system based on support vector machine (SVM) classifier. The proposed work depends on the handwriting word level, and it does not need for the character segmentation stage. In feature extraction phase, three types of features are obtained from character images which are structural features, statistical features which includes connected components feature, zoning features and global transformation which include Discrete Cosine Transform (DCT) features, histogram of oriented gradient (HOG). After the feature extraction, a multiclass SVM classification has been used in the proposed system with different kernels of linear, polynomial and RBF. The final step is the recognition which is matching the selected class by the SVM with the character ASCII and finds the desired word in the Arabic lexicon. The proposed Arabic handwriting dataset has 2913 handwriting word images. Each word has 105 images written in a different style. In handwritten word recognition system, 70% of the dataset is used for training purpose and 30% for testing and it achieved 96.31% recognition accuracy with polynomial kernel of SVM classifier. Tamen et al. [70] proposed an efficient multiple classifier system for Arabic handwritten words recognition. They used Chebyshev moments (CM) enhanced with some statistical and contour-based features (SCF) for describing word images. They considered statistical features based on local information at the edges of the forms in order to be able to discriminate word images with globally similar shapes. For the recognition improvement purpose, they used a combination of classifiers like multilayer perceptron (MLP), support vector machine (SVM) and Extreme Learning Machine (ELM) classifiers. Each classifier is trained with the two sets of features CM and SCF separately resulting in six classifiers, namely, MLP_CM, MLP_SCF, SVM_CM, SVM_SCF, ELM_CM and ELM_SCF. In order to achieve good results, a combination is first attempted by training a MLP network and an

ELM one. Then they proposed several combination rules of the six resulting classifiers which are: max-rule (MR), average-rule (AR), general weighted average-rule (GWAR), intra-class weighted average-rule (ICWAR), max-average combination rule (MACR), average-max combination rule (ARCM), Borda Count, Dempster–Shafer combination rule. A second level of the combination is tested with three of the rules, namely the GWAR, the Borda Count and the Dempster–Shafer rules. It consists in estimating the AR rule upon these rules. The average is computed using the results of these combination rules. The system is evaluated on the IFN/ENIT database and compared to some well-known systems for Arabic handwriting recognition. The proposed system is able to achieve a global recognition rate of 96.82% for the considered dataset.

Moubtahij et al. [44] presented an off-line handwritten Arabic script recognition system based on Hidden Markov Models Toolkit (HTK). In this approach, *a priori* segmentation of words is needed. The proposed system is divided into two parts. The first part of the writing recognition system is the preprocessing phase that prepares the data which serve to introduce and extract a set of simple statistical features of a sliding window along that text line from the right to left. Extracted feature vector is composed of four characteristics: intensity, normalized gray level, horizontal gray level derivative and vertical gray level derivative. The output of this step is a sequence of 80 different components for each window (20 components per each characteristic). Part two is performed inside Hidden Markov Model Toolkit (HTK). It assembles the feature vectors with the corresponding transcriptions of the handwritten text. In the recognition phase, the concatenation of characters to form words is modeled by simple lexical models. Each word is modeled by a stochastic finite-state automaton (SFSA) which represents all possible concatenations of individual characters that may compose the word. By embedding the character HMMs into the edges of this automaton, a lexical HMM is obtained. These HMMs estimate the word conditional probability. The proposed system is applied to an “Arabic-Numbers” data corpus, which contains 47 words and 1905 sentences. These sentences are written by five different people. The evaluation experiments were derived from a set of 86 images for the test and a set of 1819 images lines for training. They worked with the parameters like 20 cells per each characteristic, 8 states per HMM, Filter Gaussian. Word recognition accuracy of 80.33% has been achieved by them.

6.1.1.2 Chinese Chinese is spoken by the Han majority and many other ethnic groups in China. Approximately 1.2 billion people (around 16% of the world’s population) speak some form of Chinese as their first language. The written form of the standard language is based on the logograms known as Chinese characters. Wang et al. [77] intro-

duced a new feature extraction method using pulse coupled neural network (PCNN) for isolated word recognition. The PCNN is used to extract the time series and entropy series from the spectrogram of words to construct the PCNN coefficients. DTW (Dynamic Time Warping) technology is the classic algorithm of realizing template matching in a speech recognition system which accomplishes the task of isolated word recognition. The distance matrix is calculated by using a DTW matching method between the reference library and test library. Finally, the identification results are obtained according to the judgment logic. All of the speech signals are Mandarin Chinese recorded by a female at a sampling frequency of 8-kHz, using 8-bit coding. The original one is labeled “a” in the corresponding filename used to generate the reference library. The other one is labeled “b” in the filename served as the test speech to produce the test library. Statistical results show that the resulted speech recognition rate remains around 96.67%. The proposed word recognition system can load the original and test speech signals automatically and can save the stage results step by step. Zhang et al. [81] adopted a two-step way to process Chinese place name recognition. In the first step, they acquired the statistical knowledge of the Chinese character in place names, and the last character used in a Chinese place name to trigger place name recognition processing. The N-gram model is used to find the Chinese place name candidates from the corpus. And in the second step maximum entropy model is used for fine recognizing processing. As to the features of the maximum entropy model, they added hierarchical network of concept (HNC) semantic concept features in order to express linguistic knowledge in detail. They took 12 kinds of classification as maximum entropy model features. The test set contains 17,825 Chinese place names in close-test and 10,065 Chinese place names in the open-test. The recognition result value is about 84% when the features just take account of the word and its part of speech. And they added HNC semantic concept features to the maximum entropy model which utilizes word and its part of speech, and the additional features help to improve recognition effect.

6.1.1.3 Dutch Dutch is a West Germanic language that is spoken by around 23 million people, including the population of the Netherlands and about sixty percent that of Belgium. It is the third most widely spoken Germanic language, after English and German. Waard [76] proposed a method to optimize the parameters of the substitution costs of a minimal edit distance. With the appropriate substitution costs, it is possible to compute the minimal edit distance between a string and a sequence of real vectors. A discriminative error criterion is proposed for the optimization with a gradient search algorithm. The minimal edit-distance operator is characterized by the allowed substitutions and the costs

of the substitutions. The search for the optimal transformation path which leads to the minimal edit distance implies a dynamic warping mechanism. The estimation of the derivative of the error with respect to the parameters of the substitution costs the error criterion can be minimized. The proposed method was tested in a basic experiment, a string-to-string matching problem. Strings were derived from explicit character segmentation and character recognition of Dutch city names from live mail. The lexicon used in the basic experiment contains 2599 words (200 alias words). The learning set contains 2015 images with 404 different words, and the test set contains 1376 images with 318 different words. On an average, the proposed system achieved an accuracy of 45.0 and 52.0% on simple and complex substitutions, respectively. Shridhar et al. [63] presented a context-driven Dutch city name recognition based on a feature driven word-matching algorithm. This algorithm uses the segmentation-recognition strategy augmented by dynamic programming to select the best match from the available lexicon. Local chain-code histograms of the character contour are used as a feature vector. The impact of varying the lexicon size and completeness is studied. They used the lexicon directed algorithm. The proposed method used Dutch city name database which contains 2940 distinct names. Including the variations, the city name lexicon has 3088 entries. The test set was composed of 9726 city name images. These test images have 1161 distinct words. To create the lexicons for testing the completeness, they picked at random 10, 25, 50, 75, and 100% from the 1161 distinct words, ranked from most to least frequent word. They achieved a 60.0% accuracy with a 2.0% error rate on Dutch city name recognition.

6.1.1.4 Japanese Japanese consists of two scripts (referred to as Kana) called Hiragana and Katakana, which are two versions of the same set of sounds in the language. Hiragana and Katakana consist of a little less than 50 “letters,” which are actually simplified Chinese characters adopted to form a phonetic script. Chinese characters, called Kanji in Japanese, are also heavily used in the Japanese writing. Most of the words in the Japanese written language are written in Kanji (nouns, verbs, adjectives). There exists over 40,000 Kanji where about 2000 represent over 95% of characters actually used in writing the text. There are no spaces in Japanese so Kanji is necessary in distinguishing between separate words within a sentence. Maruyama and Nakano [41] proposed a recognition method for cursive Japanese words written in Latin characters by the integration of two classifiers, i.e., pattern matching based on directional features and Hidden Markov Model (HMM). Pattern matching based on directional features is known as an effective method in handwritten kanji recognition. Four patterns emphasizing four directions (vertical, horizontal, left slant, right slant) at every block are formed. The similarity of a pattern is

the average of similarities in four directional patterns. The similarities are calculated for templates of all classes, and the best candidate classes are selected. Three candidates are outputted in descending order of similarities for each segmented pattern. Candidates for patterns generated from the same group are merged and sorted in descending order of similarities. The templates used in the matching are generated from averaging features of learning samples. The total number of templates for 52 classes (Latin characters, lower and uppercase) is 5200. In feature extraction phase for HMM, by scanning the normalized pattern vertically from top to bottom, four features are extracted. After that, from bottom to top, a similar operation extracts the other four features. The eight dimensional vectors are used as the feature vector. Since the operation is repeated for each pixel on the abscissa, 16 vectors are generated. An HMM is constructed for each character class. Therefore, 52 HMMs are constructed corresponding to A–Z and a–z. Though there are lower and uppercases, the recognition results are merged after recognition. Using HMMs constructed by learning, recognition is executed. For an input pattern, the probability of each HMM is calculated. The recognition result is the class corresponding to the highest probability outputted from the HMMs. Three candidates are outputted for each segmented pattern, merged, and then sorted in descending order of the probabilities in the same group. The method integrates pattern matching and HMM classifiers using duplicated candidates in these classifiers and orders of classifiers to improve the word recognition rate combining their results. Though the first rank recognition rate using only the pattern matching is 56.8% and that using only HMM is 59.2%, the first rank recognition rate has improved to 68.4% by the integration and the cumulative recognition rate among the ten best candidates is 92.5%.

6.1.1.5 Latin/Roman Latin or Roman script is a set of graphic signs (script) based on the letters of the classical Latin alphabet containing the 26 most widespread letters. The script is either called Roman script or Latin script, in reference to its origin in ancient Rome. Latin script is the most widely adopted writing systems in the world. Latin script is used as the standard method of writing in most Western and central European languages, as well as in many languages in other parts of the world. Roy et al. [57] presented a novel verification approach toward improvement of handwriting recognition systems using a word hypothesis rescoring scheme by Deep Belief Networks (DBNs). They employed the Marti-Bunke feature to represent the binary word images in this system. Using the sliding window technique, the text image is represented by a sequence of column-wise local feature vectors. The feature consists of a set of nine features including geometrical and contour gradient information. The baseline recognition system is built on two distinct

neural networks. One popular verification approach in text recognition system is to recognize individual characters in a word. Character segmentation/boundary detection in cursive word image is difficult. So, they considered the (Hidden Markov Model) HMM-based forced alignment approach in this framework. This approach is widely used for providing consistent and accurate character segmentations. The forced alignment using the Viterbi algorithm finds the most probable boundaries for the given sequence of character units. Using the alignment algorithm, they obtained the character segments of a given word hypothesis. Then, a verification approach using a DBN classifier is performed for each character segments. A comparative evaluation of DBNs and MLPs (multilayer perceptrons) on the Rimes (Latin) word recognition dataset has been conducted in this experiment. The database consists of 59,203 Latin script word images divided into three subsets: 44,197 images for training, 7542 for validation, and 7464 for testing. In this work, they used the segmented word images and the reduced test dictionary of size 1612 words. They obtained the best accuracy at 91.84% by integrating DBN-based verification approach. The gain is 1.12% from using only BLSTM (Bidirectional Long Short-Term Memory) neural network. Using MLP based verification scheme the re-ranking score obtained is 91.03%. The results obtained show that the verification approach using DBNs outperforms that of MLP systems.

Al-Boeridi and Ahmad [3] developed off-line Handwriting Recognition (OHR) for Malaysian bank cheques written in the Malay language. The proposed system is comprised of three components, namely a character recognition system (CRS), hybrid decision system (HDS) and lexical word classification system (LWC). The CRS in this system was implemented using two individual classifiers, namely an Adaptive multilayer feed-forward back-propagation neural network (ANN) and support vector machine (SVM). Both statistical and geometrical extraction techniques have been applied to this approach. The results of the experiments show that ANN yields better recognition accuracy than SVM. The CRS achieved the best feature extraction with the ink crossing (InC) and profiles feature set. The HDS is a hybrid of two classifiers ANN and SVM, and its main function is to recognize the characters from the input data arrays. HDS finds the classifier accuracy for each character independently and does not require a training stage for the data, as it acquires a trained system from the single classifiers. The HDS inputs all the available test data, whereby it processes two arrays of character accuracy of each of the ANN and SVM recognizers and then decides which output to choose for each character case. After the CRS is executed, the final step in the HDS is that of word recognition. Word recognition step essentially relies on the Lexical Word Classification (LWC) method, which was applied twice in this study. Initially, it was applied in the single classifier directly after operating

the CRS (regardless the HDS method), and after that, it was applied after the HDS (multiple classifiers procedure). The HDS method used in this proposed system increases the accuracy of ANN–SVM output to 98.53% recognition, using the MCW database, whereby 11,890 characters were correctly classified out of 12,067 total test characters. The use of multiple classifier procedure yielded a slight improvement in word recognition, with an overall word recognition accuracy of 98.7%, when using the LWC system.

Tay et al. [71] developed an approach to combine Neural Network (NN) and Hidden Markov Models (HMM) for solving handwritten word recognition problem. Geometrical features are extracted from each frame image. The segmentation graph may contain letters hypotheses that are junks, the NN needs to be trained to tackle such problem. They introduce discriminate training, where in this process, it carefully selects letter hypotheses that are influential in causing recognition errors as junk examples, and then to retrain the NN with extra output neurons for handling junk class. Results on three databases, namely IRONOFF, SRTP and AWS, are presented and show the superiority of the hybrid recognizer compared to baseline recognizer, which is using discrete HMM. The IRONOFF contains a total of 36,396 isolated French word images from a 196-word lexicon. The off-line handwriting signals are sampled with a spatial resolution of 300 dots per inch (DPI), with 8 bits per pixel (256 gray level). A subset of the IRONOFF-196, which consists of only French cheque-word, is named IRONOFF-Cheque. IRONOFF-Cheque has only 30 word lexicons, and has 4481 test images. SRTP-Cheque database is collected from the real postal cheques by SRTP, the research arm of the French post office. It consists of 26 word lexicons with 27,638 training images and 7884 test images. The last database AWS-1334 consists of 1374 word lexicons with a total of 4035 words which is extracted from a text that belongs to the LOB (Lancaster-Oslo-Bergen) corpus and was written by single scripiter. Finally, they show that the hybrid recognizer can be bootstrapped automatically from the discrete HMM recognizer, and significantly improve its recognition accuracy by going through several training stages and achieves a 96.1% recognition rate. Naik and Patel [45] have proposed off-line Roman handwritten word recognition system using best structural features for feature extraction. In this system, input is a scanned bitmap image of the handwritten word. Various preprocessing steps are performed on the input image which includes skew detection and correction, slant detection and correction, baselines estimation and skeletonization. After preprocessing step all the features are extracted from the skeleton of an image. They have fused various features like loops, junction points, lines and endpoints to get better efficiency. For each word in a dataset, these features are calculated and stored in a feature vector to be used for classification. In this work, Euclidean distance classifier is

used for the classification task. It uses a Euclidean distance formula and outputs a single matching word. They have collected 800 samples of words from 25 different writers of 30 district names of Karnataka. It is an attempt toward automating postal services in Karnataka.

6.1.1.6 Mongolian The Mongolian language is a kind of cursive alphabetic script and has a very special writing system which is mainly used in the Inner Mongolian Autonomous Region, Heilongjiang, Jilin, Liaoning and Xinjiang provinces of China. Its writing order is vertical from top to bottom and the column order is from left to right. All letters of one Mongolian word are conglutinated together to form a vertical backbone, which makes the segmentation of Mongolian word very difficult. Liu et al. [40] presented a novel recognition method based on convolutional neural network model for Mongolian words recognition (MWRCNN) and position maps for online handwritten Mongolian words. Firstly, the input Mongolian word is converted into multiple artificial data with different scales and positions using aspect ratio and position maps based on 64×64 image data and these data have 15 different styles. In addition, several hybrid data sets are obtained by analyzing the structure of position maps. Secondly, two feature combination methods, namely, MWRCNN with n branches (MWRCNN_FC_N) and feature combination based on MWRCNN with one branch (MWRCNN_FC_ONE), are proposed to extract significant features. MWRCNN_FC_N has n different input branches and combines n features from different branches in the contact layer. However, in the architecture of MWRCNN_FC_ONE only one branch is the same as that of MWRCNN_FC_N. The other $n - 1$ kind of features are extracted from the first fully connected layer of $n - 1$ well-trained MWRCNNs with $n - 1$ corresponding transformed data. All features of different branches are combined into a branch in contact layer. Thirdly, four classifier combination methods, multi-column MWRCNN based on position maps (MCMWRCNN_PM), multi-column MWRCNN based on a different aspect ratio (MCMWRCNN_DAR), multi-column MWRCNN based on multiple feature combination (MCMWRCNN_MFC) and multi-column MWRCNN based on all above (MCMWRCNN_ALL), are presented to enhance the recognition performance using multi-column deep neural network (MCDNN) method which averages individual predictions of each classifier. The online handwritten Mongolian word database named MRG-OHMW is composed of 946 classes, each class with approximately 300 samples produced by 300 writers. 250 samples per class are selected for training, and the remaining 50 samples per class for evaluating the recognition performance. Experimental results show that the proposed methods achieved the word-level recognition rate of 92.22% with data transformation, 92.60% with multiple feature combination and 93.24% with

multiple classifier combination, respectively, far better than the benchmarking recognition accuracy 91.20%.

6.1.1.7 Persian The Persian alphabet or Perso-Arabic alphabet is a writing system used in the Persian language. The Persian alphabet contains 32 letters. This script shares many features with other systems based on the Arabic script. Therefore, a Farsi/Persian word recognizer can also be used for recognition of Arabic words. The Farsi text is inherently cursive both in handwritten and printed forms and is written horizontally from right to left. Dehghan et al. [12] have proposed a holistic system for the recognition of handwritten Farsi/Arabic words. In the holistic approach, a word is treated and identified as an entity. Discrete hidden Markov model (HMM) is chosen as the recognition engine in which a separate HMM is used for each word class. The histogram of chain-code directions of the image strips, scanned from right to left by a sliding window, is used as feature vectors. The Kohonen self-organizing feature map (SOFM) is used for preserving the neighborhood information and also smoothing the observation probability distribution. The database contains more than 17,000 images of 198 city names of Iran. A subset of 60% of images was randomly chosen for building the training data set and the remaining images were used as the testing data set. The Kohonen SOFM clustering program was used to construct a codebook from chain-code histogram feature vectors extracted from vertical frames of the training data set. For each class c , the best right-left HMM was chosen from the trained HMMs using Baum–Welch algorithm. The proposed system achieved top choice accuracy of 65.0% without rejection. Imani et al. [22] proposed an off-line recognition system for Farsi handwritten words. Two types of gradient features (directional and intensity gradient features) were extracted from a sliding vertical stripe which sweeps across a word image. The intensity feature represents the number of white pixels within each sliding window cell. The feature vector extracted from each stripe is then coded using Kohonen self-organization vector quantization to obtain a codebook with 49 symbols. After generating the codebook, a given feature vector is mapped to a symbol from 1 to 49. Thus, each word image is now identified by an observation sequence which is given as input to the hidden Markov models (HMM) classifier. The number of choosing states in HMM is proportioned to the word length. In this proposed system, a right-to-left HMM is employed. To evaluate the performance of the proposed method, FARSA dataset has been used in which 198 word classes were used. The experimental results show that the proposed system, applying directional gradient features, has achieved the recognition rate of 69.07%.

6.1.1.8 Thai Thai script is used to write the Thai, Southern Thai and other languages in Thailand. It has 44 con-

sonant letters, 15 vowel symbols that combine into at least 28 vowel forms, and four tone diacritics. Vichianchai [74] proposed the approach of Thai-word segmentation through Thai-writing structure matching. Indeed, the writing structure was originated from the words stored in the 1999 Royal Institute Dictionary and Thai-writing levels. After finding Thai-writing structure as required, there would be the deletion of the repeated words in order to gain the smallest number of the structures. After that, leftover structures would be used for the word segmentation. The documents used for the performance test of word segmentation contained a variety of data, in terms of writing patterns used for communicating with the readers under the identical understanding. In fact, the documents used in this research were concerned with newspapers, articles, Buddhism, encyclopedia, laws, non-fiction, the Royal Family's news, interviewing, and general news. These papers contained a variety of words use. According to the performance test of the word-segmentation process, it was observed that the word-segmentation accuracy was 94.0%.

6.1.1.9 Uyghur Uyghur language is Turkish language used in the Xinjiang Uyghur autonomous region in China. Uyghur characters are written in a cursive style from right to left and for it no upper or lower case exists. Its alphabet contains 32 characters, each has between two and four shapes (128 characters) and the choice of which shape to use depends on the position of the letter within its word. Ibrayim and Hamdulla [21] proposed an approach for online handwritten Uyghur word recognition using segmentation-based techniques. This handwritten word recognition system uses a lexicon of candidate strings to provide context for a word image in the handwritten word recognition process. The word recognition algorithms take two inputs: the word image and a lexicon representing possible hypotheses for the word image. The goal is to assign a matching score to each lexicon entry or to select the best lexicon entry among the set. This approach is referred to as a lexicon-driven approach because an optimal segmentation is generated for each string in the lexicon. In segmentation stage, the input handwritten word image is initially segmented into primitive segments, then construct the candidate segmentation network. A word image is segmented into sub-images called primitives. The segmentation processes use these steps such as removing delayed strokes, shape analysis of the stroke trajectory, reconstructing delayed strokes and combining adjacent fragments. In matching stage, the sequence of primitive segments matches with the character string of the lexicon. This is a dynamic programming problem, which minimizes an edit distance, and the result largely depends on the cost defined for segment-to-character match. In this system, they used two match functions: adding match function and the normalizing match function. In the test, they

used two kinds of distance measure, namely the adding edit distance and normalized edit distance. The database used by them for testing consists of 1460 words collected from different people using a Han Wang writing tablet. As a result, the performance for lexicons of size 10,100, 500 and 1000 are 93.17, 70.33, 59.79, 51.20% and 94.85, 79.75, 74.42, 62.19% for adding minimum edit distance and normalizing edit distance, respectively.

6.1.2 Work on bi-scripts

Roy et al. [56] proposed a scheme for word-wise handwritten script identification of bi-script documents written in Persian and Roman scripts. In the proposed scheme, simple but fast computable set of 12 features based on fractal dimension, position of small component, topology, etc., are used. They have used the fractal dimension of the full image, its contour, upper half and lower half of the contour of the image constituting 4 features based on the fractal dimension. Three features based on the position of the small components in respect to the word image have been used. This feature is selected on the basis that, compared to Roman the number of characters having multiple dots in upper or lower part is very common in Persian script. They have used 5 features based on the topology of the image which are: area of the loop, maximum length of the horizontal black run and its position, the statistical models of horizontal and vertical black run length. A set of classifiers such as multilayer perceptron (MLP) neural network, support vector machine (SVM), k -nearest neighbor (k -NN) and modified quadratic discriminant function (MQDF) are employed for script identification. They have used a database of 5000 handwritten words (2577 Persian and 2423 Roman handwritten words) for their experiments. Out of them 4000 (2000 each) samples are used for training of the proposed system and the rest are used for testing. The best result among all these classifiers was 99.20% obtained using SVM with Gaussian Kernel. To the best of authors knowledge, this is the first of its kind as no work on word-wise script identification for handwritten Persian and Roman scripts is available in the literature. Eynard and Emptoz [15] presented an Italic/Roman word type recognition system without a priori knowledge of the characters' font. This method aims at analyzing old documents in which character segmentation is not trivial. Therefore, this approach segments the document into words and analyzes the text per word. To define the word style, they combined three criteria which are based on the visual differences between a word and a slanted version of the same word. These criteria are defined based on features computed from the vertical projection profile of the word. The results show a ratio of 100% recognition of Italic words and 97.2% of Roman words.

6.2 Indic scripts

Script wise work related to word recognition of Indic scripts is summarized in this section:

6.2.1 Work on single script

6.2.1.1 Bengali Bengali is the national language of Bangladesh while it is also the state language of the Indian state of West Bengal. It is written from left to right. The Bengali character set is divided into 21 vowels, 36 consonants and modifiers. The vowels themselves can be divided into dependent and independent vowels. It is recognizable, as are other Brahmic scripts, by a distinctive horizontal line running along the tops of the letters that link them together which is known as “*Matra*.” Due to the structure of the Bengali script, the characters frequently overlap and hard to segment, especially when the writing is cursive. So, Adak et al. [2] developed an approach to deal with off-line handwritten word recognition of Bengali script in which a better approach is to send the whole word to a suitable recognizer. The proposed method employed a hybrid model in which convolutional neural network (CNN) is integrated with a Recurrent Model. CNN works as an unsupervised feature vector extractor. CNN behaves like a transformation function, which takes input image sequence frames and produces a feature vector as output. This feature vector is then fed into a recurrent neural network for sequence modeling. In the recurrent net, LSTM (Long Short-Term Memory) block is used as a hidden unit. The Connectionist Temporal Classification (CTC) layer resides at the end of this model, i.e., top of the recurrent net. The objective of this CTC layer is sequence labeling by determining a probability for character (or ortho-syllable) recognition. It can also eliminate ambiguity due to “blank” or a null labeling facility. They have tested this method on three datasets: (a) a publicly available dataset, which contains 17,091 words, (b) a new dataset generated by their own research group which contains 107,550 words in which there are further three datasets containing 47,200 words (high frequency word dataset), 30,600 words (Paragraph dataset containing all basic characters), 29,750 word images (Paragraph dataset containing most of the conjunct characters) and (c) an unconstrained dataset which is comprised of 5223 words. The obtained accuracies on datasets (a), (b) and (c) are 85.42, 86.96 and 70.67%, respectively.

Chowdhury et al. [10] proposed a framework to recognize online handwritten Bangla words considering different writing styles based on fuzzy features. The proposed system consists of four main modules: segmentation, fuzzy feature extraction, learning phase and recognition phase.

The input word is drawn using a mouse on the screen or on a drawing pad. This data is collected at the time ordered sequence of coordinates. After dividing the input data into segments the next step is to determine the fuzzy features of each segment, which consist of a determination of the universe of discourse, relative position and geometric feature of a segment. The universe of discourse of a character is the smallest rectangular area where the character fits completely. In the training phase, the entire system is managed using a database. In learning phase, characters, numerals and vowel signs are learned. After segmentation and feature extraction, the linguistic terms of fuzzy features are stored in the database. In the recognition phase, for the recognition of the drawn word three steps are needed to be accomplished. In the first step, all the segments of input data are categorized into a number of groups. All the segments of a single character belong to a single segment group. In the second step, features of group segments are compared with a database. For a group, such characters in the database are searched where the total number of segments in the character is equal to the total number of segments in the group. In the third step, after recognizing all characters they are aggregated sequentially and the desired word is formed. The proposed system is tested with 500 words of varying length and writing styles which were collected from a total of 10 participants. The achieved recognition accuracy is 77.0%.

Bhowmik et al. [4] presented a holistic word recognition technique for the recognition of handwritten Bangla words. In a holistic approach, the features of the entire word image are used for the recognition purpose, i.e., the individual word is treated as an individual class. In this work, a set of elliptical features is extracted from handwritten word images to represent them in the feature space. Elliptical features, considering an entire word image is computed by two ways: In the first approach, four feature values are designed by considering three concentric ellipses drawn on a word image. In the second approach, from the outermost ellipse, drawn on the word image, first 5 features are estimated. These are (i) number of foreground pixels on the boundary of the ellipse, (ii) number of foreground pixels along the axis parallel to X -axis, (iii) number of foreground pixels along the axis parallel to the Y -axis, (iv) ratio of foreground pixels and background pixels inside the ellipse and (v) ratio of foreground pixels inside and outside of the ellipse (not exceeding the minimum boundary box of the word). Then the region inside the outermost ellipse is divided into four sub-regions depending on the center and two focus points. Finally, number of foreground pixels inside these 4 sub-regions is computed. Thus, in total, 9 feature values are computed from the outermost ellipse. Therefore, from each word image, 13 global feature values are extracted. To get

the local information, a word image is divided into 4 small subparts depending on the center of the ellipse and from each subpart same feature value, as mentioned earlier, are computed. Therefore, in total 65 (i.e., 5 (4 sub-sections and 1 whole word) $\times 13$) elliptical features are computed from a particular word image. Then, a comparison among 5 well-known classifiers such as naive Bayes, bagging, dagging, multilayer perceptron (MLP) and support vector machine (SVM) is carried out in terms of their accuracies to select the suitable classifier for evaluating the present work. Based on that, finally, MLP classifier is chosen for the recognition task. For the proposed work, the database contains 20 different word classes (i.e., city names) with 51 samples per class, i.e., total 1020 words. For the recognition task a three-fold cross-validation method is used. Each fold contains 680 word images for training and 340 for testing. The proposed system achieved a recognition accuracy of 85.88%.

6.2.1.2 Devanagari Devanagari script is closely related to the Nandinagari script commonly found in numerous ancient manuscripts of South India. This script has 47 primary characters, of which 14 are vowels and 33 are consonants. It is written from left to right. The Devanagari script is used for other various languages, making it one of the most used and adopted writing systems in the world. Shaw et al. [61] proposed hidden Markov model (HMM) based approach to recognition of handwritten Devanagari words. The histogram of chain-code directions in the image strips, scanned from left to right by a sliding window, is used as the feature vector. This scheme is based on a holistic approach, which extracts global features from an image, thus reducing the overhead of segmentation. A handwritten word image is assumed to be a string of several image frame primitives which are in fact the states of the proposed HMM and are found using a certain mixture distribution. One HMM is constructed for each word. To classify an unknown word image, its class conditional probability for each HMM is computed. The class that gives the highest such probability is finally selected. The database consists of 39,700 samples of handwritten words collected from 436 different writers which are in fact the names of 100 towns in India. It consists of 22,500 training word images and 17,200 test word images. The correct classification rate, misclassification rate and rejection rate obtained on the test set are 80.2, 16.3 and 3.5%, respectively.

Patil and Ansari [51] proposed online handwritten Devanagari word recognition system in which the input image is drawn on a smartphone. Gesture class is used to capture gesture which user will draw on the screen. Feature extraction of the input image is done by android technology. Then features stored in the file are recognized by hidden

Markov model (HMM). For experimental analysis they divided words into two categories. In category 1 word does not contain any modifier while category 2 word consists of lower and upper modifier. They obtained the recognition accuracy of 96.0 and 94.6% within the top 5 and 10 candidates from the lexicon, respectively, for category 1 words and 95.7 and 94% recognition accuracy within the top 5 and 10 candidates from the lexicon, respectively, for category 2 words. They tested their application to 50 and 100 words by different writers. They achieved recognition accuracy of 96 and 94% on 50 and 100 words, respectively. They concluded that the recognition accuracy of the word up to 3 characters is greater than up to 5 characters. Kumar [38] proposed a segmentation-based approach for Devanagari hand-printed word recognition. In this approach, the location of shirorekha (headline) is estimated and finally it is removed. This disassociate the “consonants” and complete “matra” or a part of “matra” above and below the “shirorekha.” A headline removal algorithm used in case of individual characters has been extended for word level headline estimation and removal. The dataset of consonants with lower “matra” is created separately. If such “matra” exists in the lower region and is not touching any consonants, then these are recognized using dedicated classifier. The individual characters in the middle region are extracted using connected component techniques based on chain code, i.e., boundary tracking method. The small components developed near “shirorekha,” during its removal are ignored. To deal with touching characters, the technique of width estimate and recognize is followed. In this strategy a tentative character sized window is moved over the merged characters and prior trained classifier is used to recognize the character. If the various components are not recognized in a single pass, in that case multiple passes are used to find the exact class of a character. In recognition strategy, the individual components available as a segmentation process are passed to the appropriate classifier to know its identity, i.e., class on the basis of its properties known as features. Multilayer perceptron (MLP) classifier is used for the recognition purpose. This algorithm has been tested on 3600 words collected from more than 200 writers. The segmentation rate for words having a headline nearly straight is 86.6 and 76.6% for two characters and six characters’ words, respectively. The recognition rate is 93.4 and 93.5% for two character and six character words, respectively. The overall recognition rate is 80.8 and 72.0% for two character and six character words, respectively.

Shaw et al. [62] presented their study on fusion of the information at feature and classifier levels for improved performance of off-line handwritten Devanagari word recognition. They considered two features, i.e., directional distance distribution (DDD) and

Gradient-Structural-Concavity (GSC) features along with multi-class support vector machine (SVM) classifiers. The gradient part of the GSC features detects local information about stroke shape on a small scale. Its structural part extracts useful information about stroke trajectories at longer distances. The concavity features detect stroke relationships spanning across the entire image. In the present work, they used float values of these features and varied the grid size of their computation to investigate the respective recognition performances. They got $32 \times m \times n$ dimensional GSC feature vector (with variable grid size $m \times n$). DDD features include both global and local shape information of the input pattern. In DDD, distances in 8 directions are computed for each pixel of the input image. Both black and white pixels are considered in this distance computation. Thus, two sets of 8 bytes, called W set and B set, are used to store these distance values for each pixel of the binary input image. The 8 bytes of W are used to encode the distances of a white pixel to its nearest black pixels in the 8 directions and the corresponding bytes of B is filled with zeros for a white pixel. Similarly, for a black pixel, the set B is used to store the distances to its nearest white pixels in 8 directions and the corresponding W is filled with the value zero. A special case occurs when a ray from a black or a white pixel hits the boundary without meeting any white or black pixel. In such cases, different strategies like the mirror and circular tiling options may be applied. In the proposed study, they used DDD features with the circular tiling option in which a ray encountering the boundary moves to the opposite side and continues in the same direction. They got $16 \times m \times n$ (a grid of size $m \times n$ over the image and in each block of this grid, they obtain the average values of all the 16 WB (White-Black) coding) dimensional DDD feature vector. They studied various combinations of DDD features along with one or more features from the GSC feature set. In the combination of GSC and DDD features, they used SVM classifier for handwritten unconstrained word recognition task. They trained two SVMs, one with GSC features and the other with DDD features. The normalized output vectors of these two SVMs, corresponding to each training sample, are concatenated to form the training set for another SVM with input vector size $2N$ and output vector size N . This SVM used in the second stage combines the output of the two SVMs of the first stage. Experimental results are obtained on a large handwritten Devanagari word sample image database of 100 Indian town names. It includes 22,500 training word images and 17,200 test word images and achieved maximum recognition accuracy of 88.75%.

6.2.1.3 Gujarati The Gujarati script was adapted from the ancient Nagari script for writing in the Gujarati language.

Gujarati is a language from the Indo-Aryan family of languages, used by more than 50 million people in the Indian states of Gujarat, Maharashtra, Rajasthan, Karnataka and Madhya Pradesh, and also in some countries like Bangladesh, Fiji, Kenya, Malawi, Mauritius, Oman. Gujarati words do not have a header line or shirorekha, as Bangla or Hindi words. The character set of Gujarati language comprises of 35 consonants, 13 vowels and 6 signs, 13 dependent vowel signs, 4 additional vowels for Sanskrit, 9 digits and 1 currency sign. The consonants can be combined with the vowels and can form compound characters. Patel and Desai [48] described an important phase of Optical Character Recognition (OCR) namely zone identification of Gujarati words. Zone identification is used to extract modifiers in the upper side of the base character called upper zone containing upper modifiers and lower side of the base character called lower zone containing lower modifiers. Detection of upper zone and lower zone will lead to detection of the middle zone which contains most of the basic characters, conjunct characters and few modifiers, if any, as part of the middle zone. It is the first attempt for Gujarati handwritten word zone identification. It is assumed that the input words images are preprocessed, noise free binary images. The approach describes a process based on the distance transform for identifying various zones for handwritten Gujarati words. This proposed approach firstly generates the Euclidean distance, transformed image for the input binary image of the word. Next, it computes the horizontal sum of each row for this transformed image and determine the approximate mid row position of the word. Next, process each upper and lower half from the mid-point, to find out two consecutive minimum sum values starting from the middle and moving toward the top and bottom side, respectively. If the difference between the positions of consecutive minimum values in respective half parts is less than 15% of the word width, then the modifiers are very small. In this case considers the previous minimum position as the separation point of the mean line or baseline. A larger difference between two consecutive minimum value positions indicates, a prominent trough, providing the position of the separation point for mean line or base line. If the minimum position is nearly equal to 1 or 2 for the top half, it indicates that no upper modifier exists in the word. If the minimum position is nearly equal to word width for the bottom half, it indicates that no lower modifiers exist in the word. Then separate each zone based on the boundary positions determined and display the ones whichever is/are present in the word. They have tested this algorithm for 250 different handwritten Gujarati words. The recognition accuracies achieved for upper, middle and lower zones are 75.2, 75.2 and 83.6%, respectively.

6.2.1.4 Gurmukhi Gurmukhi is a popular language not only in India but also in the world as it is the 14th most widely spoken language in the world. It is the script used for writing the Punjabi language. Gurmukhi has 3 vowel carriers, 38 consonants, 9 vowels, 3 half vowels, and 3 half characters. In Gurmukhi script, most of the work is reported on character recognition level. For example, Sharma and Jhaji [60] have achieved maximum recognition accuracy of 72.5 and 72.0%, respectively, with k -NN and SVM for isolated off-line handwritten Gurmukhi characters. Kumar et al. [34] have also presented a character recognition using principal component analysis. They have explored k -NN and SVM classifiers for off-line handwritten character recognition [29–33]. Kumar et al. [35, 36] have also presented efficient feature extraction techniques based on curvature features and a hierarchical technique for off-line handwritten Gurmukhi character recognition. Dhiman and Lehal [14] presented a comparative performance analysis for Gurmukhi optical character recognition (OCR) at the word level. The most important step in any OCR system is to extract the features of images. For feature extraction, word images have been scanned and these images are machine-printed images. Discrete cosine transform (DCT) and Gabor filter have been used to extract the features. A Gabor filter is selective to both spatial frequencies as well as orientation frequency so sometimes called as a kind of local narrow band pass filter. The Gabor filter provides 189 features for scanned images. DCT is the member of a family of sinusoidal unitary transforms, which encodes the significant details or energy or frequency of the image in a few coefficients very efficiently. These transformed coefficients are used as features of the sample image. DCT provides 100 features of scanned images in zig-zag method. Next comes the classification, which assigns the input features of the stored pattern and compares it to find out the best matches. They have used k -NN (k -nearest neighbor) classifier for recognition of word images. In the testing stage, k -NN classifier is popular and simplest classifier. First, the system is trained with some samples. It simply stores training samples with its label. For prediction of a sample, its distance is computed from the training sample. After computing distance, the k closest training samples are kept, where k is a fixed integer having a value $k \geq 1$. After that a label is searched. This searched label is a most common label among all those samples, which is the prediction for the test sample. In this approach, $k = 3$ and 5 is chosen for the minimum distance. To train the classifier of Gurmukhi OCR, 50 different classes with 30–35 samples of each class, i.e., 1600 samples have been taken. 750 samples have been used to test the system. Using a Gabor filter, k -NN classifier provides 92.62% of correctness while with DCT, k -NN provides 96.99% of accuracy.

6.2.1.5 Kannada Kannada is the official language of Karnataka state. It is the 27th most spoken language in the world. Kannada has its own script derived from Bramhi script. Kannada script has a set of 49 characters. They are classified into three categories: Swara (vowels), vyanjana (consonants), and yogavahakas. There are 13 vowels, 34 consonants and 2 yogavahakas. Patel et al. [50] have focused on the Kannada off-line handwritten word recognition, which contains 5 stages. In data acquisition stage, handwritten words are collected to make the dataset which is further scanned and stored in a standard format. In preprocessing stage, image is enhanced to improve the recognition rate, which includes gray scale conversion, binarization, noise removal, skew correction and corner point detection. In feature extraction stage, all the essential features of the scanned image are extracted and they have used Locality Preserving Projections (LPP) for the feature extraction. LPP is a linear dimensionality reduction algorithm and it overcomes the disadvantages of Principal Component Analysis (PCA). LPP mainly focuses on the local structure of the image. But PCA focuses on the global structure. Given the set of data points and local similarity matrix, they find the optimal projections by solving the minimization problem, i.e., comparing each data point to its neighborhood. Next comes the classification and recognition phase. During the training phase, images are grouped into separate classes and each image is labeled. In the testing phase, features of the test image are compared with the trained images. Matching word images are displayed later, which is done using classifiers. Here they have used support vector machines (SVM) classifier and the result is compared with the k -means classifier. Post-processing is the last step in their proposed off-line Kannada handwritten word recognition system. The dataset contains handwritten words of 30 district and 174 Taluk names of Karnataka state written by 50 people. The dataset is prepared by different people with the geographical area of the Karnataka state. The average recognition rate achieved on dataset with proposed SVM classifier is 85.0% which is better than k -means classifier's average recognition rate of 83.0%.

Gowda et al. [18] proposed an off-line Kannada handwritten word recognition system using locality preserving projections (LPP) method for the feature extraction. Support vector machine (SVM) is employed to classify the images. After the classification of the images according to their density, the accurate image of the input word is recognized and displayed. The dataset for the proposed system consists of 30 districts of Karnataka and those words sampling are taken from the 20 people. The system identified the words and the recognition rate was observed 80% on average for all

the possible variations. Patel and Reddy [49] addressed the impact of the grid based approach in off-line handwritten Kannada word recognition. The proposed method divides the input word into four grids. For each grid, subspace learning approach, i.e., principal component analysis (PCA) (eigen vectors) is applied for better representation. These eigenvectors represent the local information of the input word which in turns helps in understanding the variations in the word image. The task of classification is to use the feature vectors provided by the feature extractor to assign the object to a category. In the proposed work, Euclidean distance measure is used to classify the input word image. They evaluated the proposed method with a dataset containing handwritten words pertaining to 28 district names of Karnataka state, each having 40 different samples and achieved the recognition accuracy of 68.57%. The experiment result revealed that the proposed grid based approach with subspace learning approach outperforms standard PCA approach.

6.2.1.6 Maithali The Maithili is a Brahmi-based script that is mostly used in the state of north Bihar in India and Nepal. The Maithili is an officially recognized language in India and the second most commonly spoken language in Nepal. Ranjan and Dubey [53] proposed Maithili dialect Isolated Word Recognition (IWR) system to recognize a spoken word by a person through a microphone or some other devices based on hidden Markov model (HMM). In the preprocessing step of IWR system, environmental noise is eliminated on the original speech signal to extract the feature. This step involves pre-emphasis, framing and windowing of signal. The aim of feature extraction is to extract relevant information from each speech frame, as a feature vector. In this work, they computed Mel-frequency cepstral coefficients (MFCC) as features by first performing a standard Fourier analysis, then converting the power spectrum to a Mel-scale and computing the inverse Fourier transform of the logarithm of that frequency spectrum. In this work, they have used two models, namely acoustic model and language model for automatic speech recognition (ASR) system. In between two models, a phonetic dictionary is also used for the matching of the ASR system. In HMM-based speech recognition system, it is assumed that the sequence of observed vectors corresponding to each word is generated by a Markov chain. The HMM model has been applied on observational sequence for each word and then by selecting, the maximum of the word is recognized. For experiment, they have used eight Maithili vowels, each vowel having two words. The speech signals are recorded using the audacity software repeatedly five times for every word. The performance is measured by the fivefold cross-validation process on the recorded data sets of eighty speech utterances. They

have computed the minimum misclassification rate at a particular value of the number of hidden state, the number of frequencies extracted from each frame.

6.2.1.7 Malayalam Malayalam is the principal language of the South Indian State of Kerala. It belongs to the southern group of Dravidian Languages. It is spoken by over 50 million people. The Malayalam character set consists of 95 characters consisting of 13 vowels, 36 consonants, 5 chillu, 4 consonant signs, 12 vowel signs, numbers and rest contributing to anuswaram, etc. Kumar and Chandran [28] developed a handwritten Malayalam word recognition system in two phases: (i) to recognize handwritten Malayalam character (ii) to develop the Malayalam word recognition system using Neural Networks with back-propagation algorithm. The feature extraction method used in the proposed work is direction feature extraction. The line segments that would be determined in each character image were categorized into four types: (1) vertical lines, (2) horizontal lines, (3) right diagonal and (4) left diagonal. Aside from these four line representations, the technique also located intersection points between each type of line. In order to provide an input vector to the neural network the character representation was broken down into a number of windows of equal size (zoning) whereby the number, length and types of lines present in each window was determined. An input vector containing 9 floating-point values were defined as: 1. The presence of horizontal lines, 2. Total length of horizontal lines, 3. The presence of right diagonal lines, 4. Total length of right diagonal lines, 5. The presence of vertical lines, 6. The total length of vertical lines, 7. The presence of left diagonal lines, 8. The total length of left diagonal lines and 9. The presence of intersection points. Each of the 10 feature vector values of the 9 zones are obtained. So, a total of 95 values is found. This will constitute the input vector to the neural network. The multilayer perceptron (MLP) network implemented for the purpose of this project is composed of 3 layers, one input, one hidden and one output. The input layer consists of 90 neurons which receive pixel binary data from a 15×12 symbol pixel matrix. The hidden layer consists of neurons whose number is decided on the basis of optimal results on a trial and error basis. The output layer is composed of neurons corresponding to each Malayalam characters. Back propagation is a systematic method of training multilayer artificial neural network. The back-propagation algorithm works in much the same way as the name suggests: after propagating an input through the network, the error is calculated and the error is propagated back through the network while the weights are adjusted in order to make the error smaller. After obtaining a recognized character, characters are grouped to form a word.

Then a database is created for a specific number of words and the written word is compared with the word stored in the table. If the word is found, then the appropriate Unicode of the characters are retrieved.

6.2.1.8 Oriya The Odia script developed from the Kalinga script, one of the many descendents of the Brahmi script of ancient India. It is the official language of Odisha, and the second official language of Jharkhand. Odia is an Indo-Aryan language spoken by about 40 million people, mainly in the Indian state of Odisha, and also in parts of West Bengal, Jharkhand, Chhattisgarh and Andhra Pradesh. The alphabet of the modern Oriya script consists of 11 vowels and 41 consonants. These characters are called basic characters. The writing style in the script is from left to right. The concept of upper/lower case is absent in Oriya script. Sahu and Mati [58] proposed an isolated odia word recognition using two approaches. The first process is the feature extraction model and second one is the feature matching model. For feature extraction model, they have used MFCC (Mel-Frequency Cepstral Coefficients) and for recognition, they have used DTW (Dynamic Time Warping). In this process, they have first recorded the speech with 16KHZ sampling frequency; then they did some preprocessing steps like calculating the energy, spectrogram and power spectral density (PSD). PSD is calculated to know the amount of power in a speech signal. After calculating the power, they analyzed the spectrogram of the input speech signal using wideband spectrogram and narrowband spectrogram. Spectrogram can be used to identify the spoken word phonetically and to analyze the various calls of human. After calculating the preprocessing steps they did a MFCC analysis of the spoken word. The process of obtaining Mel-Cepstral Coefficients involved the use of a Mel-scale filter bank. The Mel-scale is a logarithmic scale resembling the way that the human ear perceives sound. In feature matching stage, the calculated features of a word are compared with the help of the database to calculate the exact spoken word. DTW algorithm is implemented to calculate the least distance between features of word utterance and reference templates. Corresponding to least value among a calculated score with each template, the word is detected. The system is trained by saving templates of five separate words. Efficiency in detecting isolated words is 100% for two syllable words compared with one syllable word. From the results above, they inferred that the DTW distance between identical words is less than 150 and between different words is more than 300. So setting the threshold of 200 can easily filter the word uttered by the user from the other words whose templates are saved in the training phase.

Mohanty and Swain [42] have developed Oriya isolated word (speech) recognition system based on hidden Markov

model (HMM) which can easily convert isolated answers uttered by visually impaired students to isolated text. In this research, the feature vectors consist of 13-dimensional Mel-frequency cepstral coefficients (MFCC) which are used in both training and testing stage. Each of the feature vectors carries significant information about the spectrum and the amount of energy in different frequency bands of a waveform at a given point in time. MFCC are calculated by passing the Oriya speech input signal through the sampling step, typical values are 16 kHz sampling rate with 16 Bit quantization. They have used HMM for recognizing Oriya isolated word sequence uttered by a human while answering the questions. HMM models spoken utterances as the outputs of finite-state machines (FSMs). In this paper, they experimented using left-to-right model of HMM which allows states to transit to themselves or to successive states but restricts transition to earlier states. The most probable word is given the observation sequence that can be computed by taking the product of the two probabilities for each word, and choosing the word for which this product is greatest. Isolated words are collected by considering the closed ended objective type of questions of five categories. In this research study, they have considered four questions of each type (category) which carries maximum four options. As a result of which they have 60 isolated words belonging to total 20 questions for five types of objective questions. In their study, they considered 30 speakers for training purpose which results 1800 isolated words for training the Oriya isolated recognition system. The system was tested on 5 speakers on both training and testing data and achieved word recognition accuracy of 76.23 and 58.86% on training and testing data, respectively. In this research study, the main objective is to provide fairness of assessment procedures for visual impaired candidates to answer their examination papers without pen and keyboard as well as in a more natural way.

6.2.1.9 Tamil The Tamil script is used by Tamils and Tamil speakers in India, Sri Lanka, Malayasia, Singapore and elsewhere to write the Tamil language. The Tamil script has 12 vowels, 18 consonants and one special character. This script is written from left to right. Thadchanamoorthy et al. [72] developed a Tamil off-line city name dataset and proposed a scheme for recognition. They consider a city name string as a word and the recognition problem is treated as lexicon-driven word recognition. In the proposed method, binarized city names are pre-segmented into primitives (individual character or its parts). Primitive components of each city name are then merged into possible characters to get the best city name using dynamic programming. For merging, the total likelihood of characters is used as the objective function and character's likelihood is computed based on the modified quadratic discriminant function (MQDF),

where direction features are applied. A dataset of 265 Tamil city names (109 Tamil Nadu city names and 156 Sri Lankan city and town names) is developed. The total number of city name classes is 265 and the number of samples for each country name is 100. In total there were 26,500 samples collected from all the classes. They obtained 99.90% reliability from their proposed system when error and rejection rates are 0.08 and 18.67%, respectively.

6.2.1.10 Telugu Telugu script is used to write the Telugu language, a Dravidian language spoken in the South Indian states of Andhra Pradesh and Telangana as well as several other neighboring states. Telugu uses 18 vowels, each of which has both an independent form and a diacritic form used with consonants to create syllables. The language makes a distinction between short and long vowels. Rasagna et al. [54] proposed a document level OCR which incorporates information from the entire document to reduce word error rates. First, the document images are preprocessed and segmented at the word level to generate the clusters. They used a run length segmentation algorithm (RLSA). For rapid clustering they used locality sensitive hashing (LSH) to create clusters for every word from the document images. An index is built by hashing word level features of document images. They employed a combination of scalar, profile, structural and transform domain feature extraction methods. Scalar features include the number of ascenders, descenders and the aspect ratio. The profile and structural features include: projection profiles, background to ink transitions and upper and lower word profiles. Fixed length description of the features is obtained by computing lower order coefficients of a discrete Fourier transform (DFT). The key idea in the hashing technique is to hash words using several hash functions so as to ensure that, for each function, the probability of collision is much higher for words which are more similar than for those which are dissimilar. The words obtained in a cluster are marked in the documents with unique cluster numbers. The query process is repeated for every unmarked word from the document images. Thus, all words from the document images are clustered into groups of similar words. In experimental dataset with 19,789 words, 98.1% of the words were correctly clustered. The first step is to use an OCR to recognize all the words in a cluster separately. The second step involves using the clusters to improve the recognition results which contain two different techniques: Character majority voting (CMV) in which the candidates for each symbol position are chosen by selecting that character which is in the majority at each position. If no candidate

is in the majority, the existing candidate is not replaced. The second technique is dynamic time warping (DTW) which utilizes dynamic programming to align the symbols. All matching characters of two words are aligned and the unmatched characters are identified as possible errors in word w of the recognizer. The unmatched characters are candidates to replace the erroneous characters in a word. Alignment of a word with all other words of the cluster gives a group of possible character replacements for errors. The probability of the possible correction is computed using maximum likelihood—in this case it is just the count of each candidate for that position divided by the total number of candidates at that position. If there is only one candidate at a position the probability is set to zero. This procedure is repeated to correct every word of a cluster. They experimented with two different datasets. The first experiment is done on a synthetic dataset obtained by generating images of words from the text and degrading them. This dataset has 5000 clusters. Each cluster has 20 images of the same word with different font sizes, and resolution. The second experiment is done with a real scanned Telugu book and examines the entire process, including clustering and improvements in accuracy. The average accuracy for all words which appear at least twice in the book is 70.37% for the OCR and 77.74 and 79.12% for the CMV and DTW methods, respectively.

6.2.1.11 Bi-scripts Most of the countries use bi-script documents. This is because every country uses its own national language and foreign language. Therefore, bilingual document with one language being the foreign language and other being the national language is very common. Postal documents are a good example of such bilingual/script documents. Dhandra et al. [13] developed an automatic technique for script identification at word level based on morphological reconstruction for two printed bilingual documents of Kannada and Devanagari containing English numerals (printed and handwritten). In this technique, the feature extractor consists of two stages. In the first stage, shape (eccentricity, aspect ratio) and directional stroke features (horizontal and vertical) are extracted based on morphological erosion and opening by reconstruction using the line structuring elements in vertical and horizontal directions. The average height of all the connected components of an image is used to threshold the length of the structuring element. In the second stage, average pixel densities of the resulting images are computed and the k -nearest neighbor (k -NN) classifier is used to classify the word image. For experimentation, 400 document pages obtained from various magazines, newspapers, books and other such docu-

ments containing variable font styles and sizes. They created a first dataset of 1850 word images by segmentation in which Kannada, Devanagari words are 750 each and 350 are Roman numerals. Another data set of 250 handwritten numerals of Kannada and Roman (each 125) is used by obtaining handwritten document pages from 50 writers. With these datasets the average script identification results of k -NN classifier are 96.10, 98.61, 94.2, 92.89 and 98.53% of printed Kannada words and Roman numerals, printed Devanagari words and Roman numerals, printed Kannada, Devanagari words and Roman numerals, printed Kannada words and handwritten English numerals, printed Devanagari words and handwritten Roman numerals, respectively. Further, they conducted a third set of experiments on 150 word images to test the sensitivity of the algorithm toward different font sizes and styles. These words are first created in different fonts using DTP (Desktop publishing software) packages, and then printed on a laser printer. It is noticed that, script identification accuracy achieved by third data set is consistent. For word-wise script identification of numerals and text words in bilingual documents, this work is first of its kind.

Rani et al. [52] developed a zone based approach based on Gabor filters for word level script identification. The Gabor features are extracted from the normalized image and from different zones of the image at each level. Then the script of Gurmukhi words, Roman words and Numerals has been identified by using these features and SVM classifier with different kernel functions. The testing of this scheme has been done on 11,400 words which contains 5212 Gurmukhi words, 4288 Roman words, 1900 Roman numerals and obtained 92.87% accuracy with linear SVM, 93.28% accuracy with polynomial kernel SVM and 99.39% accuracy with the RBF kernel function of SVM classifier. To the best of authors knowledge, this is the first work which identifies Roman words and Numerals from Gurmukhi script. Zinjore and Ramteke [82] proposed a simple and efficient algorithm used for identification of Devanagari (Marathi) script and extraction of Roman words from a printed bilingual document, which is helpful for developing bilingual optical character recognition (used in Post office, Bank, School, Railways, etc.). For identification and removal of the Devanagari script, morphological approach is used for obtaining the bounding box. Morphological thinning operation is applied to feature extraction. Header-line pixel counts and the inter character gap is used as a feature for script identification and a heuristic approach is used for classification. For extraction of Roman words, the closed neighbors to Bounding boxes (BB) are joined and two BB, that are on the same text line

in the image are grouped if the distance between them is less than the specified threshold. The proposed algorithm has been tested on a dataset of 10 documents with varying font size of 77 lines and 474 words and achieved a recognition accuracy of 85.95%.

Ghosh and Roy [16] presented a comparative study of three feature extraction approaches for online handwritten word recognition of two major Indic scripts, namely, Bengali and Devanagari using a hidden Markov model (HMM). First approach uses features extracted from the whole stroke without local zone division after segmenting the word into its basic strokes. The other two approaches consider the segmentation of a word into its basic strokes and a local zone wise analysis of each online stroke. For each basic stroke of a word, directional and structural features (DS), zone wise structural and directional features (ZSD), zone wise slopes of dominant points (ZSDP) features are extracted. In directional and structural features, five different features, namely, writing direction, slope, curvature, curliness and standard deviation of x and y coordinates are extracted in the locality of each point. In zone wise structural and directional features, the bounded box of each basic stroke is divided into a number of local zones of rows and columns. Similar features as described in DS approach are extracted in each of these local zones. In Zone wise slopes of dominant points, like the ZSD approach, each online stroke information is divided into a number of local zones of rows and columns. Here, for each basic stroke, dominant points are calculated in each of these zones and features corresponding to dominant points are used for recognition. Next, slope angles between consecutive dominant points are calculated separately for the portion of the trajectory lying in each zone. After that, stroke wise feature values are used to generate the features of the entire word. They applied hidden Markov model (HMM) based on stochastic sequential classifier for recognizing online characters. The HMM is used because of its capability to model sequential dependencies. Each character model is an HMM, which models the horizontal succession of feature vectors representing this character. Similarly, each vowel modifier model is an HMM. For data collection, a total of 100 writers, 50 native Bengali speakers and 50 native Hindi speakers belonging to different age groups contributed handwriting samples. A total of 350 different words each of Bengali and Devanagari script have been considered for training and testing the proposed system. The training and testing data are in 3:2 ratios. A total of 1000 words were used for each script in the lexicon. From the comparative study of the word recognition results, they have noted that the dominant point based local feature extraction provides

best accuracies for both Bengali and Devanagari scripts. They have obtained 90.23 and 93.82% accuracy for Bengali and Devanagari scripts, respectively. They have also tested their system without performing stroke segmentation on each word. It has been noted that the recognition accuracies are reduced, especially for Bengali script, compared to segmentation-based approach.

Acharyya et al. [1] developed a holistic word recognition technique for handwritten documents using a multi-layer perceptron (MLP) classifier. The holistic paradigm in handwritten word recognition treats the word as a single, indivisible entity and attempts to recognize words from their overall shape, as opposed to recognizing the individual characters comprising the word. Longest-run based holistic features are computed from the whole word image. These features are computed in four directions; row wise (east), column-wise (north) and along the directions of two major diagonals (northeast and northwest). Next word images are hierarchically partitioned into vertical segments for extraction of the local features. This hierarchical partitioning process continues up to depth-5, having 4 feature values in each segment resulting in total 252 feature values. The features, computed hierarchically on each word image are fed to an MLP based classifier, for identification of an image in one of the 12 categories of word images. To test and evaluate the proposed work the dataset CMATERdb1.2.1 is used which consists of 50 handwritten document images written in Bangla script mixed with English words. They have extracted the English words from the document pages and then total 12 word classes consisting of 291 highly frequent image sets are used for the learning purpose. The accuracy of the technique is evaluated using a threefold cross-validation method. The best-case and average-case performances of the technique to said data set is 89.9 and 83.24%, respectively.

Roy and Pal [55] proposed an automatic scheme for word-wise identification of handwritten Roman and Oriya scripts for Indian postal automation. In the proposed scheme, at first, document skewness is corrected. Next, using a piecewise horizontal projection, the document is segmented into lines and by vertical histogram the lines into words. After that by applying the water reservoir concept they computed the busy-zone of a word. A busy-zone of a word is the region of the word where maximum parts of its characters lie. Here, at first all top and bottom reservoirs are detected from the characters of a word. They calculated the average height of all the top and bottom reservoirs. For all top and bottom reservoirs whose height is less than 1.25 times the average reservoir height, they filled them by black pixel. Also, all

the loops are filled up with black pixel before computing the busy-zone. After filling the reservoirs and loops with black pixel they computed the busy-zone height of this filled-up image. Area of top and bottom reservoirs in this busy-zone plays an important role in this script identification approach. In the Neural Network based classification of the proposed scheme, they have used mainly the features like Fractal based feature, water reservoir based feature, the presence of small component, topological features, etc. The fractal dimension is an important characteristic of the fractals because it contains information about their geometric structures. They have used the fractal dimensions of the full image, its full contour, upper contour and the lower contour of the image for the script identification purpose. In water reservoir based feature, the ratio of the area of the top reservoir to the bottom reservoir of a word image is used as a feature. They have used 6 features based on busy-zone. In small component based feature, they take all the components whose height and width is less than twice of the stroke width and compares their position with respect to the busy-zone. If such components lie completely above or below the busy-zone, then those component numbers and position is used as a feature. This feature is selected based on the characteristics of scripts. They have used 9 features based on topological features of the image such as aspect ratio (ratio of length and width), area of the loop, maximum length of the horizontal and vertical black run, the statistical models of horizontal and vertical black run. Considering all the above-mentioned features a set of 25 features is generated. They have used a multilayer perceptron neural network (MLP-NN) based scheme for identification of handwritten Oriya and English script. They have used a 2-layer NN with the number of neurons in the input and output layers as 25 and 2, respectively, since the number features are 25 and the number of possible classes in handwritten script selected for the present case is 2. They have used a database of 2500 (1200 Oriya and 1300 English) handwritten words collected from postal data as well as from some individuals. Out of these data a set of 1500 (750 Oriya and 750 English) words is used for training and the rest of the dataset is used for the purpose of testing for the experimental work. From the experimental study, they obtained 99.6% accuracy in the training data and 97.69% in the testing data. The proposed scheme is independent of text size and there is no need of any normalization of the image in the proposed technique. In Table 2, we have presented comparisons of word recognition results achieved by different researchers for various scripts with different features and classification techniques.

Table 2 Recognition results of word recognition methods

Author	Dataset	Feature extraction	Classification technique	Accuracy
Dehghan et al. [12]	More than 17,000 images of 198 city names of Iran	Kohonen self-organizing feature map based on histogram of chain-code directions	Hidden Markov model (HMM)	65.0%
Verma et al. [73]	Training and testing sets of 3106 and 317 words, respectively, taken from the CEDAR benchmark database	Modified Borda Count (MBC)	MUMLP, GUMLP, MURBF	91.0%
Su and Lu [69]	Four datasets, including three ICDAR Robust Reading Competition datasets (ICDAR 2003, ICDAR 2011 and ICDAR 2013) and Google Street View Text dataset (SVT)	HOG (Histograms of Oriented Gradients)	Two multilayer Recurrent Neural Network (RNN) model with bidirectional Long Short-Term Memory (LSTM) and Connectionist Temporal Classification (CTC)	ICDAR03: 89%, ICDAR11: 87%, ICDAR13: 90% SVT: 89%
Hafiz and Bhat [20]	IFN-ENIT Database of Arabic words: Set A: 600 word images Set B: 400 word images Set C: 200 word images	Part-of-Arabic-Word (PAW) vector	Hybrid HMM-KNN (hidden Markov model, <i>k</i> -nearest neighbor)	Set A: 82.67% Set B: 86% Set C: 94%
Al-Boeridi and Ahmad [3]	MCW database, containing 12,067 Malay language test characters	InC (Ink Crossing) and Profiles feature set, PS and CG (Pixel density and Center Gravity distance) set	Adaptive multilayer feed-forward back-propagation neural network (ANN) and support vector machine (SVM)	98.7%
Wang et al. [77]	All of the speech signals are Mandarin Chinese recorded by a female at a sampling frequency of 8-kHz, using 8-bit coding	Pulse coupled neural network (PCNN) based on Time series and entropy series from the spectrogram of words	DTW (dynamic time warping) technology	96.67%
Cheikh and Kacem [9]	Tunisian city names extracted from IFN/ENIT database (15 samples for each of the 50 city names)	TNN-DF uses structural features for the global description and the recognition of the words and Fournier descriptors (DF) for the local description of some letters only if an ambiguity exists	Three NN-MLPs (Neural Network-Multilayer perceptrons)	77.6%
Zhang et al. [81]	17,825 Chinese place names in close-test and 10065 Chinese place names in open-test	Maximum entropy model features, HNC (Hierarchical Network of Concept) semantic concept features	N-gram model	84.0%
Khemiri et al. [27]	IFN/ENIT database containing 7881 Arabic word samples	Structural features	Bayesian networks, notably Naive Bayes (NB), tree augmented naive Bayes network (TAN), horizontal and vertical hidden Markov model (VH-HMM) and dynamic Bayesian network (DBN)	VH-HMM: 90.02%
Patil and Ansari [51]	(i) 50 Devanagari words (ii) 100 Devanagari words	Android technology based on gesture class	Hidden Markov model (HMM)	(i) 96% (ii) 94%

Table 2 (continued)

Author	Dataset	Feature extraction	Classification technique	Accuracy
Adak et al. [2]	(i) A publicly available dataset, which contains 17,091 Bengali words, (ii) A new dataset generated by our research group which contains 107,550 words (iii) An unconstrained dataset which is comprised of 5223 words	Convolutional neural network (CNN)	Recurrent neural network	(i) 85.42%, (ii) 86.96% (iii) 70.67%
Chowdhury et al. [10]	500 Bengali words collected from a total of 10 participants	Fuzzy features	Features of group segments are compared with database to recognize a word	77.0%
Waard [76]	The lexicon used in the basic experiment contains 2599 Dutch city names words (200 alias words)	Minimal edit distance based on appropriate substitution costs	Dynamic warping mechanism	Accuracy of 45.0 and 52.0% on simple and complex substitutions, respectively
Kumar [38]	3600 Devanagari words	Connected component technique based on chain code, i.e. boundary tracking method	Multilayer perceptron (MLP)	Recognition rate is 80.8 and 72.0% for two character and six character words, respectively
Liu et al. [40]	Online handwritten Mongolian word database named MRG-OHWW containing 2,82,954 samples	MWRCNN_FC_N and MWRCNN_FC_ONE	Four classifier combination methods: (i) multi-column MWRCNN based on position maps (MCM-WRCNN_PM), (ii) multi-column MWRCNN based on different aspect ratio (MCMWRCNN_DAR), (iii) Multi-column MWRCNN based on multiple feature combination (MCMWRCNN_MFC) and (iv) Multi-column MWRCNN based on all above (MCMWRCNN_ALL), using Multi-column deep neural network(MCDNN) method	93.24%
Patel et al. [50]	Handwritten words of 30 district and 174 taluk names of Karnataka state written by 50 people	Locality preserving projections (LPP)	Support vector machines (SVM)	85.0%
Naik and Patel [45]	800 samples of Karnataka words	Structural features	Euclidean distance classifier	–
Gowda et al. [18]	30 districts of Karnataka taken from the 20 people	Locality preserving projections (LPP) method	Support vector machine (SVM)	80.0%
Shaw et al. [61]	39,700 Devanagari words	Histogram of chain-code directions	Hidden Markov model (HMM)	80.2%
Thadchanamoorthy et al. [72]	26,500 Tamil word samples	Modified quadratic discriminant function (MQDF) based on directional features	Dynamic programming	99.90%
Septi and Bedda [59]	48 cities of Algeria in shape of Arabian manuscript	Topological features	Network of neurons of type multilayered perceptron	90.00 and 98.33%

Table 2 (continued)

Author	Dataset	Feature extraction	Classification technique	Accuracy
Shridhar et al. [63]	9726 Dutch city names	Local chain-code histograms of the character contour	Word-matching algorithm which uses segmentation-recognition strategy augmented by dynamic programming	60.0%
Caesar et al. [7]	(i) 1280 US city names (ii) 200 German city names	All features based on a symbolic representation of the contour and the skeleton	Hidden Markov models (HMM)	(i) 91% (ii) 85%
Dhandra et al. [13]	(i) 1850 word images (Kannada, Devanagari words are 750 each and 350 are English numeral) (ii) 250 handwritten numerals of Kannada and English (each 125)	Morphological reconstruction based on shape and directional stroke features	k -Nearest neighbor classifier (k -NN)	98.61%
Rami et al. [52]	11,400 words which contains 5212 Gurmukhi words, 4288 English words, 1900 English numerals	Gabor filters	SVM	92.87, 93.28, and 99.39% with Linear-SVM, Polynomial-SVM and RBF-SVM, respectively
Zinjore and Ramteke [82]	474 words (Devanagari and Roman words)	Morphological thinning operation	Heuristic approach	85.95%
Tay et al. [71]	(i) IRONOFF contains a total of 36,396 isolated French word images from a 196-word lexicon (ii) SRTP-Cheque database consists of 26 word lexicons with 27,638 training images and 7884 test images (iii) AWS database consists of 1374 word lexicons with total of 4035 words	Geometrical features	Neural network (NN) and Hidden Markov models (HMM)	96.1%
Ranjan and Dubey [53]	Recorded Maithili data sets of eighty speech utterances	Standard Fourier analysis based on Mel-frequency cepstral coefficients (MFCC) as features	Hidden Markov model (HMM)	95.0%
Bouaziz et al. [5]	Dataset of 500 Arabic words written by four writers	Structural features	RBF-SVM	96.82%
Khaissidi et al. [26]	Large amount of handwritten Ibn Sina datasets containing Arabic documents	Histogram of oriented gradient (HOG)	Linear-SVM	68.4%
Ibrayim and Hamdulla [21]	Database containing 1460 Uyghur words collected from different people	Minimal edit distance based on cost defined for segment-to-character match	Dynamic programming	94.85%
Jayeche et al. [23]	IFN/ENIT database which consists of 946 handwritten Tunisian city names	Zernike and HU moments	Dynamic hierarchical Bayesian network	82.0%
Karim and Kadhim [24, 25]	Arabic handwriting dataset AHDB containing 2913 handwriting word images	Discrete cosine transform features (DCT) and histogram of oriented gradient (HOG)	Neural network	95.0%

Table 2 (continued)

Author	Dataset	Feature extraction	Classification technique	Accuracy
Karim and Kadhm [24, 25]	Arabic handwriting dataset AHDB containing 2913 handwriting word images	Discrete cosine transform features (DCT) and histogram of oriented gradient (HOG)	Support vector machine	96.31%
Imani et al. [22]	FARSA dataset containing 198 word classes	Kohonen self-organizing feature map (SOFM) based on gradient features	Hidden Markov model (HMM)	69.07%
Acharyya et al. [11]	The dataset CMATERdb1.2.1 is used which consists of 50 handwritten document images written in Bangla script mixed with English words	Hierarchical partitioning	Multilayer perceptron (MLP)	89.9%
Dasgupta et al. [11]	Training and testing datasets containing 5230 and 2514 English words	Arnold transform based on directional features	Multi-class support vector machine	87.19%
Tamen et al. [70]	IFN/ENIT database containing Arabic words	Chebyshev moments (CM) and statistical and contour-based features (SCF)	Combinations of classifiers like multilayer perceptron (MLP), support vector machine (SVM) and extreme learning machine (ELM)	96.82%
Ghosh and Roy [16]	350 different words each of Bengali and Devanagari script	Directional and structural features (DS), zone wise structural and directional features (ZSD), zone wise slopes of dominant points (ZSDP) features	Hidden Markov model (HMM)	90.23 and 93.82% accuracies for Bengali and Devanagari scripts, respectively
Shaw et al. [62]	22,500 training word images and 17,200 test word images	Directional distance distribution (DDD) and gradient-structural-concavity (GSC) features	Multi-class support vector machine	88.75%
Moubtahij et al. [44]	“Arabic-Numbers” data corpus, which contains 47 words and 1905 sentences written by five different people A set of 86 images for the test and A set of 1819 images lines for training	Intensity, Normalized gray level, horizontal gray level derivative and vertical gray level derivative	Hidden Markov Models Toolkit (HTK)	80.33%
Bhowmik et al. [4]	Database contains 20 different word classes with 51 samples per class, i.e., total 1020 words: 680 word images for training and 340 for testing	Elliptical features	Multilayer perceptron (MLP)	85.88%
Patel and Reddy [49]	Handwritten words comprising of 28 district names of Karnataka state, each having 40 different samples	Grid based approach with subspace learning approach, i.e., principal component analysis	Euclidean distance measures	68.57%
Singh et al. [65]	Database containing 61 words of English script written by 63 writers where each writer writes each of the words five times resulting in $(61 \times 63 \times 5 =)$ 19,215 word samples	Template formation	Template matching technique using correlation coefficient	76.4%

Table 2 (continued)

Author	Dataset	Feature extraction	Classification technique	Accuracy
Roy et al. [57]	The database consists of 59,203 Latin script word images divided into three subsets: 44,197 images for training, 7542 for validation, and 7464 for testing	Marti-Bunke features including geometrical and contour gradient information	Bidirectional long short-term memory blocks (BLSTM) as hidden layer and the connectionist temporal classification (CTC)	91.84%
Patel and Desai [48]	250 different handwritten Gujarati words	Zone identification	Euclidean distance transform	Upper zone: 75.2% Middle zone: 75.2% Lower zone: 83.6%
Vichianchai [74]	Number of Thai words in the dictionary that the program could correctly carried out word segmentation was 93,294	Thai-word segmentation	Thai-writing structure matching	94.0%
Roy et al. [56]	Database of 5000 handwritten words: 2577 Persian and 2423 Roman handwritten words	Fractal dimension, position of small component, topological features	Multilayer perceptron (MLP) neural network, support vector machine, k -nearest neighbor and modified quadratic discriminant function	99.20%
Dhiman and Lehal [14]	50 different classes with 30–35 samples of each class, i.e., 1600 Gurmukhi samples have been taken 750 samples have been used to test the system	Gabor filter provides 189 features for scanned images and discrete cosine transform (DCT) provides 100 features of scanned images in zigzag method	k -Nearest neighbor	96.99%
Kumar and Chandran [28]	Malayalam words	Direction feature extraction	Multilayer perceptron (MLP) network	–
Eynard and Emptoz [15]	Gazette of Leyde dataset containing words typeset in Italic in paragraphs typeset in Roman	Features computed from the vertical projection profile	Combination of three criteria which are based on the visual differences between a word and a slanted version of the same word	97.2% Roman words
Rasagna et al. [54]	The first experiment is done on a synthetic dataset containing 5000 clusters. Each cluster has 20 images of the same word The second experiment is done with a real scanned Telugu book	A combination of scalar, profile, structural and transform domain feature extraction methods	The first step is to use an OCR to recognize all the words in a cluster separately The second step involves using the clusters to improve the recognition results which contains two different techniques: character majority voting (CMV) and dynamic time warping (DTW)	70.37% for OCR 77.74% for CMV and 79.12% for DTW methods
Maruyama and Nakano [41]	A lexicon size of 166 and the test set contains 1312 Japanese cursive words written in Latin characters	Pattern matching based on directional features and HMM based on 16 dimensional feature vector	Pattern matching and hidden Markov model (HMM)	92.5%

Table 2 (continued)

Author	Dataset	Feature extraction	Classification technique	Accuracy
Roy and Pal [55]	A database of 2500 handwritten words containing: 1200 Oriya words 1300 English words (i) Training set contains 1500 (750 Oriya and 750 English) words and (ii) the rest of the dataset is used for the purpose of testing	A set of 25 features is generated containing: (i) Fractal based feature, (ii) Water reservoir based feature, (iii) Presence of small component, (iv) Topological features	Multilayer perceptron neural network (MLP-NN)	99.6% accuracy in the training data and 97.69% in the testing data

7 Inferences and future directions

In order to know the advances in word recognition, authors have attempted to find the relevant literature related to word recognition of text documents printed/handwritten in non-Indic and Indic scripts. This literature has been surveyed script wise and summarized it in the form of a table. This survey is helpful in finding the research gaps in area of word recognition. Various combinations of classifiers have been proposed in the literature for improving the word recognition accuracy. A word in a document may consist of more than one connected component which creates a problem in segmenting the word into characters to recognize a word. So, holistic approach is performing better than segmentation-based technique for word segmentation. But, still there is a scope for research in off-line handwritten word recognition due to the complexities of the scripts, various writing styles, etc. So in future research scholars can propose a scheme that can provide comparable and if possible better performance than already established off-line handwritten word recognition schemes. The word-segmentation accuracy can further be increased by proposing more efficient algorithms for word segmentation. They can use some features specific to the most confusing characters, to increase the recognition rate. The number of samples for training and testing can also be experimented for improving the recognition accuracy of the system. The lexicon size and completeness are important factors that clearly limit the accuracy of handwritten word recognition, so it needs to be considered. The future work also needs an efficient segmentation method that can segment each text word without any overlapping or missing parts.

Appendix 1: A quality assessment forms

Screening question

Section-1

Does the research paper refer to word recognition? Yes

Consider:

The paper includes the study of word recognition. All types of studies, i.e., case study, experimental study or research paper is included.

Section-1 is evaluated first. If the reply is positive, then proceed to Section-2.

Screening question

Section-2

Key sub-area categorization

Is the research paper focusing on word recognition? Yes

Consider:

- Is the study’s focus or main focus on word recognition or not?
- Did the study fit in any one of the sub-areas categorized? (Apparently the study motivated different categories.)

If the study’s primary focus is on word detection, proceed to section-3, else proceed to section-4.

Detailed questions

Section-3

Findings

Is there clear statement of the findings? Yes

Consider:

- Did the study mention the approach/word detection?
- Has the word detection technique reported?
- What is the corresponding transformation technique, findings, i.e., source representation?

Comparison

Was the data reported sufficient for comparative analysis? Yes

Consider:

- Are the necessary parameters for comparison discussed?
- Is the study referring to handwritten word recognition explicitly?

Detailed questions

Section-4

Findings

Did the study mention the type of word recognition? Yes

Consider:

- How well the word recognition is categorized?
- Did the study explicitly mention the type of word recognition, or is to be inferred from the study?

Appendix 2: Data items extracted from all papers

Data item	Description
Study identifier	Unique ID for the study
Bibliographic data	Author, year, title, source
Type of article	Journal article, conference article, workshop paper

Data item	Description
Study aims/context/application domain	What are the aims of the study, i.e., search focus, i.e., the research areas the paper focus on
Study design	Classification of study—feature extraction, classification, word recognition, comparative analysis, etc.
What is the word recognition technique?	It explicitly refers to the techniques used for extracting the features of word, segmentation techniques if any and classification techniques to recognize a word
How was comparison carried out?	Values of important parameters for word recognition, i.e., recall, precision, application area, scalability, portability
Subject system	How the data was collected: it refers to the subject system and its size
Data analysis	Data analysis, i.e., corresponding source representation and match detection techniques are extracted
Developer of the tool and usage	It refers to the word detection tool, developer and usage of the tool
Study findings	Major findings or conclusions from the primary study like percentage of word’s recognition accuracy
Other	Does the study explicitly refer to handwritten word recognition or printed word recognition, any other important point

References

1. Acharyya A, Rakshit S, Sarkar R, Basu S, Nasipuri M (2013) Handwritten word recognition using MLP based classifier: a holistic approach. *Int J Comput Sci Issues* 10(2):422–427
2. Adak C, Chaudhuri BB, Blumenstein M (2016) Offline cursive Bengali word recognition using CNNs with a recurrent model. In: *Proceedings of the 15th international conference on frontiers in handwriting recognition*, pp 429–434
3. Al-Boeridi ON, Ahmad SM (2015) A scalable hybrid decision system (HDS) for Roman word recognition using ANN SVM: study case on Malay word recognition. *Nat Comput Appl Forum* 26(6):1505–1513
4. Bhowmik S, Malakar S, Sarkar R, Nasipuri M (2014) Handwritten Bangla word recognition using elliptical features. In: *Proceedings of the sixth international conference on computational intelligence and communication networks*, pp 257–261
5. Bouaziz S, Mezghani A, Kanoun S (2014) Arabic handwritten word recognition with large vocabulary based on explicit segmentation. In: *Proceedings of the international conference on*

- information and communication technologies innovation and application, pp 1–4
6. Bouwhuis D, Bouma H (1979) Visual word recognition of three letter words as derived from the recognition of the constituent letters. *Percept Psychophys* 25:12–22
 7. Caesar T, Gloger JM, Kaltenmeier A, Mandler E (1994) Handwritten word recognition using statistics. In: Proceedings of the IEE European workshop on handwriting analysis and recognition: a European perspective, pp 1–5
 8. Cattell J (1886) The time taken up by cerebral operations. *Mind* 11:277–282
 9. Cheikh IB, Kacem A (2007) Neural network for the recognition of handwritten Tunisian city names. In: Proceedings of the international conference on document analysis and recognition, pp 1108–1112
 10. Chowdhury K, Alam L, Sarmin S, Arefin S, Hoque MM (2015) A fuzzy features based online handwritten Bangla word recognition framework. In: Proceedings of the 18th international conference on computer and information technology (ICCIT), pp 484–489
 11. Dasgupta J, Bhattacharya K, Chanda B (2016) A holistic approach for Off-line handwritten cursive word recognition using directional feature based on Arnold transform. *Pattern Recogn Lett* 79:73–79
 12. Dehghan M, Faez K, Ahmadi M, Shridhar M (2001) Handwritten Farsi (Arabic) word recognition: a holistic approach using discrete HMM. *Pattern Recogn Lett* 34(5):1057–1065
 13. Dhandra BV, Mallikarjun H, Hegadi R, Malemath VS (2006) Word-wise script identification from bilingual documents based on morphological reconstruction. In: Proceedings of the first international conference on digital information management, pp 389–394
 14. Dhiman S, Lehal GS (2017) Performance comparison of Gurmukhi script: k-NN classifier with DCT and Gabor filter. *Int J Adv Res Comput Sci* 8(5):762–764
 15. Eynard L, Emptoz H (2009) Italic or roman: word style recognition without a priori knowledge for old printed documents. In: Proceedings of the 10th international conference on document analysis and recognition, pp 823–827
 16. Ghosh R, Roy PP (2016) Comparison of zone-features for online Bengali and Devanagari word recognition using HMM. In: Proceedings of the 15th international conference on frontiers in handwriting recognition, pp 435–440
 17. Gough PB (1972) One second of reading. In: Kavanagh JF, Mattingly IG (eds) *Language by ear and by eye*. MIT Press, Cambridge
 18. Gowda PK, Chethan S, Harsha J, Rakesh J, Tanushree KN (2017) Offline Kannada handwritten word recognition using locality preserving projections (LPP). *Int J Innov Res Comput Commun Eng* 5(5):9955–9960
 19. Fisher DF (1975) Reading and visual search. *Memory Cogn* 3:188–196
 20. Hafiz AM, Bhat GM (2016) Arabic OCR using a novel hybrid classification scheme. *J Pattern Recognit Res* 11(1):55–60
 21. Ibrayim M, Hamdulla A (2015) On-line handwritten Uyghur word recognition using segmentation-based techniques. *Int J Signal Process Image Process Pattern Recognit* 8(6):51–60
 22. Imani Z, Ahmadyfard AR, Zohrevand A (2016) Holistic Farsi handwritten word recognition using gradient features. *J Artif Intell Data Min* 4(1):19–25
 23. Jayech K, Mahjoub M, Amara NB (2016) Arabic handwritten word recognition based on dynamic Bayesian network. *Int Arab J Inf Technol* 13(6B):1024–1031
 24. Karim A, Kadhm MS (2015) Handwriting word recognition based on neural networks. *Int J Appl Eng Res* 10(22):43120–43124
 25. Karim A, Kadhm MS (2015) Handwriting word recognition based on SVM classifier. *Int J Adv Comput Sci Appl* 6(11):64–68
 26. Khaissidi G, Elfakir Y, Mrabti M, Lakhliat Z, Chenouni D, El-yacoubi M (2016) Segmentation-free word spotting for handwritten Arabic documents. *Int J Interact Multimed Artif Intell* 4:6–10
 27. Khemiri A, Echi AK, Belaid A, Elloumi M (2016) A System for off-line Arabic handwritten word recognition based on Bayesian approach. In: Proceedings of the 15th international conference on frontiers in handwriting recognition, pp 560–565
 28. Kumar M, Chandran S (2015) Handwritten Malayalam word recognition system using neural networks. *Int J Eng Res Technol (IJERT)* 4(4):90–99
 29. 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
 30. 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, Wagnaghat (Shimla), pp 1–4
 31. 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, Wagnaghat (Shimla), pp 1–4
 32. 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
 33. 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
 34. Kumar M, Jindal MK, Sharma RK (2013) PCA based offline handwritten Gurmukhi character recognition. *Smart Comput Res* 3(5):346–357
 35. Kumar M, Sharma RK, Jindal MK (2014) Efficient feature extraction techniques for offline handwritten Gurmukhi character recognition. *Natl Acad Sci Lett* 37(4):381–391
 36. Kumar M, Sharma RK, Jindal MK (2014) A novel hierarchical technique for offline handwritten Gurmukhi character recognition. *Natl Acad Sci Lett* 37(6):567–572
 37. Kumar M, Jindal MK, Sharma RK, Jindal SR (2018) Character and numeral recognition for Non-Indic and Indic scripts: a survey. *Artif Intell Rev*. <https://doi.org/10.1007/s10462-017-9607-x>
 38. Kumar S (2016) A study for handwritten Devanagari word recognition. In: Proceedings of the international conference on communication and signal processing, pp 1009–1014
 39. Lehal GS, Singh C (2000) A Gurmukhi script recognition system. In: Proceedings of the international conference on pattern recognition (ICPR'00), pp 557–560
 40. Liu J, Ma LL, Wu J (2016) Online handwritten Mongolian word recognition using MWRCNN and position maps. In: Proceedings of the 15th international conference on frontiers in handwriting recognition, pp 60–65
 41. Maruyama K, Nakano Y (2000) Recognition method for cursive Japanese word written in latin characters. In: Proceedings of the seventh international workshop on frontiers in handwriting recognition, pp 133–142
 42. Mohanty S, Swain BK (2010) Markov model based Oriya isolated speech recognizer—an emerging solution for visually impaired students in school and public examination. In: Proceedings of the international conference on communications and technologies, pp 107–111

43. Mori S, Suen CY, Yamato K (1992) Historical review of OCR research and development. *Proc IEEE* 80(7):1029–1058
44. Moubtahij HE, Satori K, Halli A (2016) Recognition of off-line Arabic handwriting words using HMM toolkit (HTK). In: *Proceedings of the 13th international conference computer graphics, imaging and visualization*, pp 167–171
45. Naik A, Patel MS (2014) Offline English handwritten word recognizer using best feature extraction. *Int J Adv Comput Theory Eng (IJACTE)* 3(2):61–63
46. Obaidullah SM, Santosh KC, Halder C, Das N, Roy K (2017) Automatic Indic script identification from handwritten documents: page, block, line and word-level approach. *J Mach Learn Cybern (JMLC)*. <https://doi.org/10.1007/s13042-017-0702-8>
47. Oyedotun OK, Khashman A (2016) Deep learning in vision-based static hand gesture recognition. *Neural Comput Appl* 28(12):3941–3951
48. Patel C, Desai A (2011) Zone identification for Gujarati handwritten word. In: *Proceedings of the second international conference on emerging applications of information technology*, pp 194–197
49. Patel MS, Reddy SC (2014) An impact of grid based approach in offline handwritten Kannada word recognition. In: *Proceedings of the international conference on contemporary computing and informatics (IC3I)*, pp 630–633
50. Patel MS, Kumar R, Reddy SC (2015) Offline Kannada handwritten word recognition using locality preserving projection (LPP) for feature extraction. *Int J Innov Res Sci Eng Technol* 4(7):5078–5086
51. Patil P, Ansari S (2014) Online handwritten Devanagari word recognition using HMM based technique. *Int J Comput Appl* 95(17):17–21
52. Rani R, Dhir R, Lehal GS (2013) Modified gabor feature extraction method for word level script identification—experimentation with Gurumukhi and English scripts. *Int J Signal Process Image Process Pattern Recognit* 6(5):25–38
53. Ranjan R, Dubey RK (2016) Isolated word recognition using HMM for Maithili dialect. In: *Proceedings of the international conference on signal processing and communication*, pp 323–327
54. Rasagna V, Kumar A, Jawahar CV, Manmatha R (2009) Robust recognition of documents by fusing results of word clusters. In: *Proceedings of the 10th international conference on document analysis and recognition*, pp 566–570
55. Roy K, Pal U (2006) Word-wise hand-written script separation for Indian Postal automation. In: *Tenth international workshop on frontiers in handwriting recognition*, pp 1–6
56. Roy K, Alaei A, Pal U (2010) Word-wise handwritten Persian and Roman script identification. In: *Proceedings of the 12th international conference on frontiers in handwriting recognition*, pp 628–633
57. Roy PP, Chherawala Y, Cheriet M (2014) Deep-belief-network based rescoring for handwritten word recognition. In: *Proceedings of the 14th international conference on frontiers in handwriting recognition*, pp 506–511
58. Sahu AK, Mati GR (2016) Odia isolated word recognition using DTW. *Int J Eng Res Technol (IJERT)* 5(8):552–556
59. Septi M, Bedda M (2006) Contribution to the recognition of hand Arabic word based on neural network. In: *Proceedings of the international conference on information and communication technologies*, pp 1635–1639
60. Sharma DV, Jhaji P (2010) Recognition of isolated handwritten characters in Gurmukhi script. *Int J Comput Appl* 4(8):9–17
61. Shaw B, Parui SK, Shridhar M (2008) Offline handwritten Devanagari word recognition: a holistic approach based on directional chain code feature and HMM. In: *Proceedings of the international conference on information technology*, pp 203–208
62. Shaw B, Bhattacharya U, Parui SK (2015) Offline handwritten Devanagari word recognition: information fusion at feature and classifier levels. In: *Proceedings of the 3rd IAPR Asian conference on pattern recognition*, pp 720–724
63. Shridhar M, Kimura F, Truijen B, Houle GF (2002) Impact of Lexicon completeness on city name recognition. In: *Proceedings of the eighth international workshop on frontiers in handwriting recognition (IWFHR'02)*, pp 513–518
64. Singh G, Sachan M (2014) Multi-layer perceptron (MLP) neural network technique for offline handwritten Gurmukhi character recognition. In: *Proceedings of the IEEE international conference on computational intelligence and computing research*, pp 221–225
65. Singh S, Kariveda T, Gupta JD, Bhattacharya K (2015) Handwritten words recognition for legal amounts of bank cheques in English script. In: *Proceedings of the 8th international conference on advances in pattern recognition*, pp 1–5
66. Smith F (1969) Familiarity of configuration vs. discriminability of features in the visual identification of words. *Psychon Sci* 14:261–262
67. Sperling G (1963) A model for visual memory tasks. *Hum Factors* 5:19–31
68. Steinherz T, Rivlin E, Intrator N (1999) Offline cursive script word recognition—a survey. *IJDAR* 2(2–3):90–110
69. Su B, Lu S (2017) Accurate recognition of words in scenes without character segmentation using recurrent neural network. *Pattern Recogn Lett* 63:397–405
70. Tamen Z, Drias H, Boughaci D (2017) An efficient multiple classifier system for Arabic handwritten words recognition. *Pattern Recogn Lett* 93:123–132
71. Tay YH, Lallican PM, Khalid M, Gaudin CV, Knerr S (2010) An offline cursive handwritten word recognition system. In: *Proceedings of the IEEE region 10 international conference on electrical and electronic technology*, pp 519–524
72. Thadchanamoorthy S, Kodikara ND, Premaretne HL (2013) Tamil handwritten city name database development and recognition for postal automation. In: *Proceedings of the 12th international conference on document analysis and recognition*, pp 793–797
73. Verma B, Gader P, Chen W (2001) Fusion of multiple handwritten word recognition techniques. *Pattern Recogn Lett* 22(9):991–998
74. Vichianchai V (2011) Thai-word segmentation through Thai writing structure matching. In: *Proceedings of the international conference on modeling, simulation and control*, vol 10, pp 184–188
75. Vinciarelli A (2002) A survey of offline cursive word recognition. *Pattern Recogn* 35(7):1433–1446
76. Ward WPD (1995) An optimized minimal edit distance for handwritten word recognition. *Pattern Recogn Lett* 16:1091–1096
77. Wang GY, Zhang YM, Sun ML, Wang X, Zhang Y (2016) Speech signal feature parameters extraction algorithm based on PCNN for isolated word recognition. In: *Proceedings of the international conference on audio, language and image processing (ICALIP)*, pp 679–682
78. Woodworth RS (1938) *Experimental psychology*. Holt, New York
79. Zhang H, Cao X, Ho JKL, Chow TWS (2017) Object-level video advertising: an optimization framework. *IEEE Trans Ind Inform* 13(2):520–531
80. Zhang H, Li J, Ji Y, Yue H (2017) Understanding subtitles by character-level sequence-to-sequence learning. *IEEE Trans Ind Inform* 13(2):616–624
81. Zhang Q, Yuan Y, Li N, Wei X, Miao J (2009) A new way for chinese place name recognition. In: *Proceedings of the international conference on Asian language processing*, pp 129–134
82. Zinjore RS, Ramteke RJ (2015) Identification and removal of Devanagari script and extraction of roman words from printed bilingual text document. In: *IJCA proceedings on national conference on digital image and signal processing (DISP)*, pp 17–20