REVIEW PAPERS

AN ANALYTICAL STUDY OF HANDWRITTEN CHARACTER RECOGNITION

By

UJWAL SINGH VOHRA *   SHRI PRAKASH DWIVEDI **   H.L. MANDORIA ***

*PG Scholar, Department of Information Technology, G.B. Pant University of Agriculture & Technology, Pantnagar, India.
** Assistant Professor, Department of Information Technology, G.B. Pant University of Agriculture & Technology, Pantnagar, India.
*** Professor & Head, Department of Information Technology, G.B. Pant University of Agriculture & Technology, Pantnagar, India.

ABSTRACT

Handwritten Character Recognition is a crucial part of Optical Character Recognition (OCR) through which the computer understands the handwriting of individuals automatically from the image of a handwritten script. From a decade, OCR becomes the most important application of Pattern Recognition, Machine Vision and Signal Processing for the rapid growth of technology, which can be described as the Electronic or Mechanical conversion of the captured or scanned image. The image is converted into the machine encoded form that can be further used in machine translation, text to speech conversion, text mining and the storage of data. Selections of appropriate feature extraction and classification methods are the crucial factors for achieving a higher rate of recognition with greater level of accuracy for handwritten characters to accurately achieve recognition of each and every letter. Here, in this paper the authors attempt to give a more elaborative image for a comprehensive review that has been proposed to achieve a deep study of the handwritten characters recognition, and this data will be useful for the readers working in the field of handwritten character recognition.

Keywords: OCR, Machine Encoded Form, Translation, Text Mining, Feature Extraction.

INTRODUCTION

The field of pattern recognition and machine learning has become one of the broad areas for the researchers. The goal of researchers is to find out the algorithms that can solve on the problems of handwritten character recognition, to the incorporate human functions of character recognition to the computers [42]. As the handwritten character recognition is the subfield of the OCR, basically the purpose of OCR is to convert an image of typewritten or printed, and handwritten text into the understandable form so that the computer can be easily able to recognize the text patterns. Consequently, the OCR system can be applied in various domains with several applications such as: bank check processing, digital libraries, mail sorting, postal code recognition, security system, and various document indexing systems [6]. However, handwritten character recognition system development is a nontrivial task because a word can be depicted in many styles, this is because each and every person has their own way of writing or representing something that can be dissimilar to others like there are many fonts to represent the same thing, and there are many styles like bold, italic, underlined, etc. with different complex layouts. Depending on the type (manual script or printed) of writing that a system should recognize, operations to be performed and the outcomes can vary depending on the type (manual script or printed) of the writing. In this paper, the objective is to cover the handwriting character recognition system, for the captured/scanned images [34]. And also the authors are covering a concise survey of the available handwritten characters recognition Methodologies, different types of feature extraction methods and different types of classifiers that are being used to classify the characters [4].

Initially, the idea of retina scanner came up with the invention of retina scanner, which is basically an image transmission system uses a mosaic of photocells in order to recognize [11]. By the invention of Nipkow's sequential

scanner, the technology of today's reading machines and television was introduced in 1890. The OCR was basically examined to provide help to the blind people, but nowadays, the OCR is the vast field for researchers to give their opinions for the growth of development in Technology.

In the patent filing of Tauschek in Germany, 1929, the first evidence of optical character recognition system is available [13]. After this, in 1935, he was given the US Patent and the Handel was given independently in 1933. [39].

The present commercial OCRs can be classified into four generations according to their robustness, efficiency and versatility. The first generation of OCRs was employed in 1960, in which the reading of selected fonts and shapes of characters can be done. IBM 1418 was the first generation commercial OCR [36], in which the method of logical template matching was there. The second generations OCRs was able to recognize the characters, printed by machine and handwritten as well. This generation OCRs were more powerful than the first generation's OCRs. IBM 1287, was the first OCR of the second generation that was the fusion of analog and digital technology [50]. In the field of character recognition, many researchers has provided their opinions and proved with their results, some of them are as follows (Table 1):

Sethi and Chatterjee (1976) carried out Devanagari numeral recognition with the help of the presence or absence of the vertical, horizontal, right slant, left slant Based primitives. Recognition was performed. The main method used was the decision tree. After this, they have attempted to recognize constraint hand printed Devanagari characters using the same method [56].

Sinha and Mahabala (1979) performed Devanagari script recognition using syntactic method with an embedded picture language. It is a prototype context based recognition [58].

Hamanaka et al. (1993), suggested an offline character recognition method that can also be operative in the online Japanese characters recognition. The traditional method that was used previously was confined to the number of strokes and their orders. The offline character

recognition procedure removes these constraints on the basis of pattern matching or the orientation of featured patterns. With the increment in various feature extraction methods like: nonlinear shape normalization, nonlinear pattern matching and normalized cooperation, the recognition of characters can be improved. The rate of accuracy attained was 95.1% [25].

Ohhira et al. (1995), suggested that for automatically recognition of 6709 Chinese characters, plural combination of neural networks can be used. The plural combination of neural network basically contains four parts as: rough classification part, recognition part, fine classification part, and auto judgment part. The input data are classified by method of the characters density at rough and fine classification part. The multi-layered NN is used to recognize the characters at the recognition part of characters recognition and the auto judgment part judges and provide output for the values. The method attained an accuracy efficiency of 100% [38].

N. Sharma et al. (2006) used information of the contour points of the characters in order to achieve the handwritten Devanagari characters recognition. Based on 64-D, features for recognition of characters was demonstrated. They suggested the scheme on the basis of quadratic classifier in order to recognize of handwritten characters and the achieved accuracy was 98.86% and with the 11270 data set size, they attained a 80.36% recognition accuracy [33].

P. S Deshpande et al. (2007) proposed a Gaussian filter to provide a feature vector for the dimension of 200 (5x5x8), and the accuracy achieved was 94% with Support Vector Machines (SVM) as the classifier [40].

S. Arora et al. (2010) proposed a system for recognizing the offline handwritten Devanagri numerals using SVM and ANN classifiers. Here, in this classification step, the recognition takes place under two stages. In the first stage, the numerals are classified using MLP and then the unrecognized characters are fed to the second stage to be classified by SVM. The proposed method was tested on 18300 data samples and achieved 93.15 % accuracy of recognition [55].

Shailedra Kumar Shrivastava and Sanjay Gharde (2010) implemented the Binary classification approach of the Support Vector Machine and the linear kernel function is applied in SVM This linear SVM produces 99.48% overall recognition rate, which is the highest among all of the methods applied on the handwritten Devanagari numeral recognition system [51].

Brijmohan Singh et al. (2011) proposed two different methods for extracting features from the handwritten Devnagari characters, the Curvelet Transform, the Character Geometry, and the large dimensional features space which is handled by PCA and their recognition performances using two different classifiers were compared. Those are: the the k-Nearest Neighbour (k-NN) and Support Vector Machine (SVM) with Radial Basis Function (RBF) classifier. Different classification accuracy measures, such as False Positive (FP) Rate, True Positive (TP) Rate, Recall and F-Measure, and Precision, are used for the purpose. Gained Results show that the Curvelet that features with k-NN classifier executes the best, yielding accuracy as high as 93.21% [7].

Anshul Gupta et al. (2011) suggested feature vectors, which are generated with the use of moduli of fourier coefficients and the phase produces an accuracy of 98.74% by using the SVM classifier for the training of datasets for their final procedure, and three combination of fourier descriptors was used in parallel. In their character recognition network, there was 52 output classes. For upper and lower case characters, separate classifiers was used, and they have also tested the MLP and RBF neural networks to recognize the characters and on the same features, they tried SVM classifier and got 62.93% of accuracy on test data and finally chosen the SVM because it is better than MLP and RBF [5].

Munish Kumar et al. (2011) have conducted research based on k-NN classifier using diagonal features for offline handwritten recognition of the Gurumukhi character and used the samples of the offline handwritten Gurumukhi characters from 100 different writers. In their proposed work, the partitioning strategy has been tested and 3500 sample images of Gurumukhi characters were used to achieve training and testing. The overall accuracy attained was 94.12% using diagonal features and k-NN classifier [27].

Mohammad Abu Obaida et al. (2011) suggested a method that consisted of : i) Before performing the character recognition process, the classification of words to a language has been done, ii) For the separation of vertical connected text lines, an improved statistical methodology has been applied, iii) To recognize inter-word and intra-word distances and to judge vertical projection, hybrid word segmentation method was adopted. This method was tested on 173 handwritten documents. 91% accuracy was achieved while processing a large sized document and the time consumed was high than processing a single bit [32].

Malakar et al. (2012) proposed that the extraction of text lines using spiral run length smearing algorithm from the document images is the most essential step of an OCR system. In the handwritten document images skew, touching and overlapping of text lines makes the recognition process a real obstacle. The method was able to extract 87.07 % and 89.35% text lines successfully from the desired database [29].

Divakar Yadav et al. (2013) have proposed that in order to improve the recognition rate, the histogram of project on the basis of mean distance, histogram projection crossing based on pixel value, and vertical zero can be used. These features extraction methodologies are also powerful to extract the feature of distorted characters symbols. For the classification purpose, a Back Propagation Neural Network with two hidden layer is used. 90% correct recognition rate of performance is attained for printed Hindi texts, and the classifier is trained and tested for the above data [12].

Gaurav Y. Tawde et al. (2014) Obtained the recognition accuracy by their proposed method of about 60% to 70%. According to them, wavelet decomposition allows observing the image at hierarchical resolution levels [13].

Chanchal Bansal and Arif Khan, (2014) proposed an efficient method for recognition of Hindi handwritten numerals, with the use of energy and chain codes. For classification purpose, the SVM is used for classification.

The average recognition of 90.1% is achieved using four segment methods [8].

Rekha Singh, (2014) proposed and implemented a Feed Forward Back Propagation Network for character recognition system. Using this method, the improvement in performance of the feed forward back propagation network is achieved and the result attained was above 80% for the handwritten sample data for the untrained sets [43].

Arjun Singh and Kansham (2015) got 97.61% result using SVM and ANN for the Handwritten Devanagari Character recognition. The accuracy/recognition rate has been obtained for different size of the images by ANN and SVM classifiers [4].

Teena Mittal and Rajendra kumar (2015) implemented SVM for recognizing the linear, polynomial and RBF kernels that have been used. In the first phase, for the fixed number of frames of the speech signal, the best strategy was found. The recognition rate was highest using the linear kernel strategy. Next, in order to know LPCs and MFCs, the number of frames was diverse and the highest accuracy achieved was 96.8% [61].

Saurabh Farkya et al. (2015) described in their paper that when a whole document was scanned and passed to OCR, the rate of recognition slightly decreases to 96%. This happens mainly due to the overlapping of characters. According to them, the surveys on Support Vector Machines found to be extremely efficient and robust in handling large amount of features, as SVM was used for the purpose of classification [57].

1. What is Handwritten Character Recognition?

Handwritten Character Recognition is a field of Pattern matching for recognizing characters. The objective of handwritten character recognition system is to build an intelligent system that will allow successful extraction of characters from the handwritten documents, to digitalize and translate the handwritten texts into machine readable texts. Variations in handwriting style, which are completely different from different writers, especially for mixed languages is still a complex task in the area of handwritten character recognition [31]. Figure 1 Shows the block diagram for Handwritten Character Recognition.

According to the working procedure of handwritten characters recognition system, there are two broad

Author & Year	Classifier	Features	Accuracy
Sethi et al. (1977)	Decision Tree	Presence or absence of 4 basic primitives such as horizontal, vertical, right slant, and left slant.	Not defined
Hamanaka et al. (1993)	Pattern Matching of orientation of feature patterns	Enhancement in nonlinear pattern matching, nonlinear shape normalization, and the normalization- cooperated feature extraction method.	95.1%.
Ohhira et al. (1995)	plural combination of Neural Networks	Recognize 6709 Chinese characters.	100%.
N. Sharma et al. (2006)	A quadratic classifier-based scheme	Directional chain code information of contour points Based on the CH, 64-D features	98.86%, 80.36%
P.S. Deshpande et al. (2008)	SVM	Gaussian filter to give feature vector of dimensionality 200 (5x5x8).	94%
Shailedra Kumar Shrivastava et al. (2010)	linear kernel in SVM	Binary classification, highest recognition rate for handwritten Devanagari numerals.	99.48%
Anshul Gupta et al. (2011)	MLP, RBF and SVM	feature vector formed using moduli of Fourier coefficients and phase, 52 output classes	98% 62.93% 98.74%
Munish Kumar et al. (2011)	k-NN classifier	3500 sample images for handwritten Gurmukhi character recognition, partition strategy for selecting the training.	94.12%
Malakar et al. (2012)	spiral run length smearing algorithm	extraction of text lines from document images	87.09% 89.35%
Divakar Yadav et al. (2013)	Back-propagation neural network with two hidden layers	Histogram of projection based on mean distance, histogram of projection based on pixel value, and vertical zero crossing, distorted characters/symbols recognition.	90%
Chanchal Bansal et al. (2014)	SVM	Efficient method for recognition of Hindi handwritten numerals using energy and chain codes.	90.1%
Arjun Singh et al. (2015)	ANN and SVM	Handwritten Devanagari Character Recognition	97.61%
Teena Mittal et al. (2015)	SVM	Linear, polynomial and RBF kernels, fixed number of frames of the speech signals.	96.8%
Saurabh Farkya et al. (2015)	SVM	very efficient and robust in handling large amount of features	96%

Table 1. Literature Review

categories [34]:

- Online handwritten character recognition, and
- Offline handwritten character recognition,

1.1 Online Handwritten Character Recognition

Under this variety of handwritten character recognition, the characters that are in creation or when the user is writing, the recognition is achieved by the device. The hardware device used to recognize data is a digitizer tablet or PDA which is electromagnetic, touch sensitive or pressure sensitive. Successive movements of the pen are transformed on the device to a series of electronic signals, when a user forms some pattern of characters on the tablet. The obtained signals are then converted to letter codes by the transducer, which are operative within the computer and the text-processing applications for further processing.

1.2 Offline Handwritten Character Recognition

This variety of handwritten character recognition requires handwritten documents before performing the task of recognition. After which the document is scanned or captured, stored in the computer and then further processing for recognition is performed on that image [28]. It is more difficult to recognize the offline handwritten characters, because different people have different handwriting styles [26].

2. Why Handwritten Character Recognition is Needed?

In professional or nonprofessional environments like libraries, homes, offices, publishing houses, etc. thousands of documents and books are scanned on regular basis for backup and storage purposes [43]. A scanner or capturing device generally takes photograph of the original paper documents and provides the image-based scanned or captured documents as the output [8]. As the document contains one big image file, rather than the individual text characters, the major concern
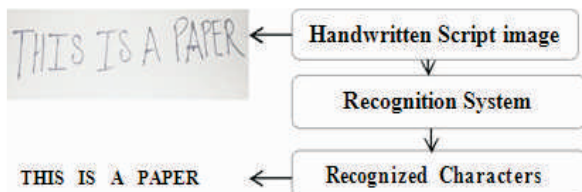

Figure 1. Handwritten character recognition

with processing and storing these kinds of large volumes of scanned or captured documents have inabilities to search for a specific name or phrase inside that file and the text cannot be highlighted, copied, or modified [6]. A large amount of time can also be saved through this process, since manual interface and human errors do not come into the play directly.

3. Working Methodologies

The working of Handwritten Character Recognition is directly proportional to the method of recognition adopted generally. Handwritten Character Recognition consists of the following stages as given in Figure 2. The output of each stage is used as the input to the next stage [34].

- Image Acquisition,
- Pre-processing,
- Segmentation,
- Feature Extraction,
- Classification, and
- Post processing,

3.1 Image Acquisition

In this phase, the image is taken through a camera or a scanner as an input. The input captured may be in the from of gray level image, color image or Binary image from the scanner or digital camera.

3.2 Preprocessing

Preprocessing involves several operations over the scanned or captured image, and so the further processing on the input image becomes suitable and comfortable for applying. Basically, the objective of
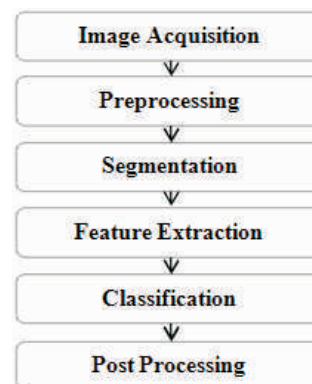

Figure 2. Stages of Handwritten Character Recognition System

preprocessing is to improve the quality of the scanned or captured input images [55]. Some of the preprocessing methods are: Gray Scale Conversion of colored image, Noise Removal, Binarization, Skew and slant detection or correction [25].

3.2.1 Noise Removal

Noise may be introduced due to any writing mistakes or disturbances during writing on a document. Sometimes, when images are scanned or captured, some scanning and capturing devices also introduce noises like bumps and gaps in line of characters, filled loops, and disconnected lines. Removing the noise is necessary for Reorganization purposes [27]. The noise can affect the image depending on the type of disturbances, to a different extent. Noise reduction techniques can be categorized in three major groups as filtering, morphological operations and Noise Modeling [53].

3.2.1.1 Filtering

The basic aim of this operation is to remove noise and diminish spurious points, which are introduced by poor sampling rate of the data acquisition device or by an uneven writing surface. Various spatial and frequency domain filters can be designed for this purpose. Filters can be designed for removing the colored background or slightly textured [35], contrast adjustments [31], and smoothing & sharpening [34], and thresholding [5] purposes from an image.

3.2.1.2 Morphological Operations

The morphological operations basically aimed to filter the document image to replace the convolution operations by the logical operations. To decompose the connected strokes [18], connect the broken strokes [6], thin the characters, smooth the contours, prune the wild points, and extract the boundaries [34], and thus various morphological operations can be designed. Therefore, the morphological operations can be successfully used to remove the noise on the document images due to erratic hand movements, as well as the low quality of paper and ink.

3.2.1.3 Noise Modeling

Some calibration techniques are applied to remove the noise, if a model for this were available. However, modeling of the noise is not possible in most of the applications. Till now, only few works have been done on modeling the noise, that is introduced by optical distortion, such as blur, [34] speckle, and skew. Nevertheless, it is possible to remove the noise to a certain degree, and assess the quality of the documents as suggested in [64].

3.2.2 Normalization

Normalization is an important phase of handwritten character recognition, Where it allows converting an image with an arbitrary size to a fixe sized image. It is generally used to reduce or eliminate the variability related to the difference in sizes and styles as much as possible [28].

3.2.3 Binarization

Binarization is performed on an image before the character recognition phase. Ideally, an input character to the recognition phase should have two tones i.e., black and white pixels (represented by 1 and 0, respectively) [35]. In image binarization, an image of up to 256 gray levels can be converted into a two toned image. The goal of this binarization step is to identify a threshold value dynamically that would help to distinguish between the image pixels, those that belong to the background and that belong to the text. The threshold value identified would be completely dependent on the nature, the contrast level, the properties of the documents, the difference of contrast between the foreground and background pixels and their concentration in the document [28]. There are various methods for binarization techniques available in image processing, out of which many researchers have found Otsu's technique to be adequate for thresholding the handwritten characters [5].

3.2.4 Skew and Slant Detection, Correction

Skew and Slant are two different important terms related to the handwritten characters. In any handwritten character recognition system, the main aim is to find out the accurate result. Skew detection/correction and Slant detection/correction are important perspectives relating to correction of the handwritten image. The amount of

skew and slant in an document may depend on different factors like writing surface and material, psychological, physical state of a person and environmental factors [2].

3.2.4.1 Skew Detection and Correction

While the scanning or capturing process of writing styles, some inaccuracies may occur, the writing may be slightly curved or tilted within the image. This can hurt the effectiveness of later phases and, therefore, skew within an image should be detected and corrected. Methods of baseline extraction for skew detection include, using the projection profile of the image, a form of nearest neighbors clustering, cross correlation method between the lines [22], and using the Hough transform. Out of them in terms of speed, the nearest neighbor method achieves the fastest method, and based on accuracy, the projection profile gives the best angle estimation accurately. After skew detection, the character or word is translated to the origin, stretched, or rotated until the baseline is horizontal and retranslated back into the display screen space.

3.2.4.2 Slant Detection and Correction

One of the important measurable factors of different handwriting styles is the slant angle between the longest stroke in a word and the vertical direction of the word. Slant detection is used to normalize all the characters to a standard form [33]. The most common method for slant estimation in a document is the calculation of the average angle of the near vertical elements. The start and end point coordinates of each line element provides the slant angle. Slant correction is the process which tries to normalize the slant of the characters in a line or paragraph to the vertical.

3.3 Segmentation

The aim of segmentation stage is to decompose an image into sub-images of individual characters. In more complex cases of handwritten text, the segmentation becomes much more difficult to process as the letters tend to be, overlapped or distorted, and connected to each other [28]. Segmentation is done to break the single text line to single word and single character from the input documents. For isolated characters, numerals and

symbols, segmentation task is not that difficult. However, more advanced techniques required to be employed for joint and complex strings. Segmentation is one important step in layout analysis and it is particularly difficult when dealing with complex layouts [34]. Layout analysis is accomplished in two stages: structural analysis is the first stage, which is concerned with segmentation of the image into blocks of document components (paragraph, row, word, etc.). Functional analysis is second one, in which, the size, location and various layout rules to describe the functional content of document components are considered. After this, by finding the textured regions in gray scale or color images, Page segmentation is implemented [31].

3.4 Feature Extraction

Feature extraction is an important methodology of Handwritten character recognition to find out the characteristics of the handwritten patterns [64]. With feature extraction, further classification for characters, recognition process can be done efficiently and effectively. A good survey for the available feature extraction methods for character recognition in presented in [1].

Feature extraction methods for documents image is categorized under three major groups.

3.4.1 Series Expansion and Global Transformation

Generally, a continuous signal contains more information that needs to be represented for the purpose of classification. By a linear combination of a series of simpler well-defined functions, a signal can be represented for discrete approximations of continuous signals, which may be true. For feature extraction, a compact encoding known as series expansion, and transformation, is provided by the coefficients of the linear combination. Under global transformation and series expansion, invariants are Deformations like translation and rotations. The following are the methods used in common transform and the series expansion for handwritten characters recognition field [35][64].

3.4.1.1 Fourier Transforms

The ability to recognize the position-shifted characters, when it observes the magnitude spectrum and ignores

the phase, and the general procedure is to choose magnitude spectrum of the measurement vector as the features in an n-dimensional Euclidean space is one of the most attractive properties of the Fourier transform [44] [68].

3.4.1.2 Gabor Transform

This is one of the variant of the windowed Fourier transform method. Generally, in this terminology, the window used for transformation is not a discrete size but defined by the Gaussian function [34].

3.4.1.3 Wavelets

It is a series expansion technique, which allow the users to represent the signal at different levels of resolutions. The segments of document image are represented by wavelet coefficients, which may correspond to letters or words for various levels of resolution. After this, these coefficients are then fed to a classifier for recognition purpose [72]. The local difference in handwriting as opposed to MRA with high resolution can be consumed by the representation in Multi Resolution Analysis (MRA) with low resolution. However, the important details for the recognition stage would be lost by the representation in low resolution.

3.4.1.4 Moments

Various moments like central moments, Legendre moments, and Zernike moments form a compact representation of the original image of the document that make the process of recognizing an object scale, invariant, rotation and translation [70],[66]. In this type of feature extraction, the original image can be completely reconstructed from the moment coefficients, so the moments are considered as a series expansion representation.

3.4.1.5 Karhunen–Loeve Expansion

Basically, this is an eigenvector analysis, which tries to reduce the dimension of the feature set by creating new features that are the linear composition of the genuine ones. In terms of information compression, Karhunen–Loeve expansion is the only ideal transformation. It is used in various pattern recognition problems such as, face recognition. At National Institute of Standards and Technology (NIST), it is also used in OCR system for form-based handprint recognition [23]. The use of Karhunen–Loeve features in CR problems is not widespread, because it requires computationally complex algorithms. However, by the increase of the computational power, it will become a realistic feature for CR systems in the next few years [71].

3.4.2 Statistical Representation

For Statistical Representation, Statistical distributions of the point, takes care of style variations to some extent for the representation of a document image [63]. It is used for reducing the dimension of the feature set providing high speed and low complexity, although this type of representation does not allow the reconstruction of the original image [39].

For character representation, the following are the major statistical features used:

3.4.2.1 Zoning

In this method, the frame containing the character is divided into various overlapping or non-overlapping zones of smaller fragments of areas. Some features in different regions or the densities of the points are analyzed and form the representation. For example, the bending point features, represent high curvature points, terminal points, and fork points [63]. Another example is contour direction features, that measure the direction of the contour of the character [21], which is generated by dividing the image array into rectangular and diagonal zones and computing the histograms of chain codes in these zones.

3.4.2.2 Crossings and Distances

A widespread statistical feature is the line segment in a specified direction, for calculating number of crossings of a contour [37]. In this, the character frame is partitioned into a set of regions in various directions and after this, by the powers of two, the black runs in each region are coded. Another study encodes number of transitions and the location from background to foreground pixels along with vertical lines throughout the word [22]. The statistical features used are the distance of line segments from a given boundary, such as the lower and upper portion of

the frame [37]. These features imply that, a horizontal threshold is established, below, above and through the center of the normalized script. The number of times a threshold is crossed by the script, it becomes the value for that feature. The main motive is to catch the ascending and descending portions of the script.

3.4.2.3 Projections

By projecting the pixel gray values onto lines in various directions, Characters can be represented. This representation creates a one-dimensional signal from a two-dimensional image, which can be further used to represent the image of characters [69],[21].

3.4.2.4 Border Transition Techniques

In this technique, it is assumed that all the characters are oriented vertically. In this, each characters are divide into four equal quadrants, the scanning and calculation of zero to one transition of each division is taken place in both the vertical and horizontal directions [63].

3.4.3 Geometrical and Topological Representation

Various local and global properties of characters can be represented by geometrical and topological features with style variations and high tolerance to distortions.

Various categories under which the topological and geometrical representations can be grouped are:

3.4.3.1 Extracting and Counting Topological Structures

A predefined structure is searched in a character or word in this group of representation. A descriptive representation forms the number or relative position of these structures within the character. Common primitive structures in this are the strokes, which make up a character. These primitives can be as simple as arcs (c) and lines (I), which are the main strokes of Latin characters and can be as complex as the curves and splines making up Chinese or Arabic characters. In online CR, a stroke is also defined by the line segments from pen-down to pen-up [9]. Successful representation of Characters and words can be done by extracting and counting the topological features such as, the maxima and minima, extreme points, openings to the right, cusps above and below a threshold, left, up, and down, loops, cross (x) points, branch (T) points, line ends (J), inflection between two

points, isolated dots, a bend between two points, direction of a stroke from a special point, symmetry of character, straight strokes between two points, ascending, descending, horizontal curves at top or bottom, middle strokes and relations among the stroke that make up a character, etc. [73], [46], [45].

3.4.3.2 Approximating and Measuring the Geometrical Properties

In many studies (e.g., [62] and [1]), by the measurement of the geometrical quantities like the ratio between height and width of the bounding box of a character, the characters can be represented. The relative distance between the last point and the last y-min, the relative vertical and horizontal distance between first and last points, distance between two points, comparative lengths between the two strokes, width of a stroke, and word length, upper and lower masses of words were analyzed. The change in the curvature or curvature is a very important characteristic measure [36]. Among many methods for measuring the curvature information [15], one is suggested to measure the local stroke direction, for measuring distribution of the directional decomposition to the directional distance of the character image. By a more convenient and compact geometrical set of features, the measured geometrical quantities can be approximated. A class of methods includes polygonal approximation of a thinned character [16]. Cubic spline representation is a more precise and expensive version of the polygonal approximation [22].

3.4.3.3 Coding

In this, one of the most widespread coding scheme is Freeman's chain code. This coding is essentially obtained by mapping the strokes of a character into a two dimensional parameter space, which is made up of codes. There are various versions of chain coding. As an example, in [74], the character frame is divided to left–right sliding window and each region is coded by the chain code.

3.4.3.4 Graphs and Trees

Characters or words are first partitioned into a set of topological primitives, such as strokes, cross points, loops,

etc. Then, these primitives are further represented using attributed or relational graphs [67]. There are two types of image representation by graphs ,the first one uses the coordinates of the character shape [14]. The second type is an abstract representation with nodes corresponding to the strokes and edges, and relationship between the strokes [65]. Trees can also be used to represent the words or characters with a set of features, which have a hierarchical relation [47][48]. In [17], from the raw data, feature extraction and selection is defined as extracting the most representative information, which minimizes the class pattern variability while enhancing between the class pattern variability. For selecting the k-best features out of N features, feature selection can be formulated as a dynamic programming problem as for a cost function like Fisher's discrimant ratio. By using principal component analysis or a NN trainer, feature selection can also be accomplished. In [29], the performance of different types of feature selection methods for characters recognition are explored and compared. Selection of features using a methodology as mentioned here most of the time supplies a suboptimal solution [42], and requires expensive computational power. Therefore, the feature selection is, mostly, done by intuition or by heuristics for a specific type of the CR application.

3.5 Characters Classification

Handwritten character recognition system broadly uses the methodology of pattern recognition, which assigns an unknown sample to a predefined class. Several techniques for handwritten character recognition are investigated by the researchers [62],[34]. Classification techniques can be classified as follows.

- Template Matching,
- Statistical Techniques,
- Neural Networks,
- Support Vector Machine (SVM) algorithms, and
- Combination Classifier.

The above approaches are not necessarily independent or disjoint from each other. Occasionally, a CR technique in one approach can also be considered to be a part of other approaches.

3.5.1 Template Matching

It is a easiest way of character recognition, based on matching the stored prototypes against the character or the word to be recognized. The degree of similarity between two vectors (group of pixels, shapes, curvature, etc.) is determined, by the matching operation. With a standard set of stored prototypes a gray-level or binary input character is compared, according to a similarity measure (e.g. Euclidean, Jaccard or Yule similarity measures, Mahalanobis, etc). A template matcher can unite multiple information sources, including match strength and k-nearest neighbor measurements from different metrics. Rate of recognition of this method is extremely sensitive to noise and image deformation. Deformable Templates and Elastic Matching are used for improved classification [18].

3.5.2 Statistical Techniques

Statistical decision theory is focused on statistical decision functions and a set of optimality criteria, to maximize the probability of the considered patterns for given model for a certain class [41]. Basically statistical techniques depend on the following assumptions:

- Gaussian is the distribution of the feature set or uniform in the worst case.
- Sufficient statistics are available for each class.
- A set of features which represents each distinct class of patterns, given collection of images is able to extract them all.

The measurements collected from n-features of all the words unit can be thought to represent the vector and n-dimensional vector space, whose coordinates correspond to the taken measurements, represent the original word unit. The various main statistical methods, applied in the OCR field are Nearest Neighbors (NN) [49],[61], Likelihood or Bayes classifier [2], Hidden Markov Modeling (HMM) [59], Clustering Analysis [37], Quadratic Classifier [33], and Fuzzy Set Reasoning.

3.5.3 Neural Networks

Human beings can recognize characters in the

documents by their learning and experience and the classification of characters problem is related to heuristic logic. Hence neural networks which are more or less heuristic in nature are highly suitable for this type of problem. Various types of neural networks are used for the handwritten character recognition classification. A computing architecture which consist the massively parallel interconnection of adaptive 'neural' processors is a neural network [20]. Compared to the classical techniques, it can perform computations at a higher rate because of its parallel nature. It can easily adapt changes in the data and learn the characteristics of input signal. Output from one node is supplied to another one and the final decision depends on the complex interaction of all nodes. Deferent type of approaches exist for training of neural networks viz. Boltzman, error correction, competitive learning, and Hebbian. They cover binary and continuous valued input, as well as supervised and unsupervised learning. Neural network architectures can be categorized as, feed-forward and feedback (recurrent) networks. The multilayer perceptron (MLP) of the feed forward networks is the most usual neural networks used in the character recognition systems and one of the interesting properties of MLP is that, in addition to classifying an input pattern, they also provide a confidence in the classification [3]. These confidence values may be used for rejecting a test pattern in case of doubt. MLP is applied by [49][61]. A detailed comparison of various NN classifiers is made by M. Egmont-Petersen et al. [24]. They have shown that, Feed-forward, perceptron higher order network, Neuro-fuzzy system are better suited for character recognition. K. Y. Rajput and Sangeetha Mishra [19] used back propagation type NN classifier. In English, genetic algorithm based feature selection and classification along with combination of NN and Fuzzy logic has been reported [59], [56].

3.5.4 Support Vector Machine Classifier

It is a supervised type of classification method in the discipline of statistical learning theory, which has been successfully applied in the domains of bioinformatics and pattern recognition applications like text, face and voice recognition. In order to maximize the distance of the separating boundary between the two or more classes, SVM tries to maximize the distance of the separating planes from each of the feature vector whether the feature vector belongs to class c_1, c_2 and so on. In SVM, the optimization criteria is the width of the margin between the classes, i.e., empty area around the decision boundary specified by the distance to the nearest training pattern [3]. These patterns, called support vectors, finally describe a classification function. By maximizing the margin, their number is minimized. Many researchers applied SVM successfully viz., Arjun Singh and Kansham [4], C. V. Jawahar et al. [10], Sandhya Arora et al. [58], P. S Deshpande et al. [40], Chanchal Bansal and Arif khan[8], Teena Mittal and Rajendra kumar[60], Saurabh Farkya et al. [57], Umapada Pal et al. [62] for characters recogiton .

3.5.5 Combination Classifier

All the classification methods have their own merit and demerits. To solve a given complex classification problem, multiple classifiers are combined together in order to provide better results. Different classifiers trained on the same data may not only vary in their global performances, but they also may show strong local differences. Every classifier may have its own zone, where it performs best in the feature space. Some classifiers such as, neural networks show different results with various initializations due to the randomness inherent in the training procedure. Instead of choosing the best network and discarding the others, one can unite various networks, thereby taking advantage of all the attempts to learn from the data [3].

3.5.5.1 Selection and Training of Individual Classifiers

Especially, if the individual classifiers are largely independent, a classifier combination is useful. Different types of resampling techniques such as, bootstrapping and rotation can be used to produce artificial variations, to increase and improve the rate of classification, if this is not already guaranteed by the use of different training sets earlier [3].

3.5.5.2 Combiner

Combiners can be distinguished from others by the

method of adapting, training and the need for the output of individual classifiers. Various combiners are used like sum (or average), voting, and the fixed type which does not require training while others are trainable. The trainable combiners provide better improvements in the system as compared to the non-trainable or static ones. On the basis of input patterns, some of the combination schemes are adaptive in the sense that the combiner can analyze (or weighs) the decisions for different types of classifiers. In contrast, all the input patterns are treated same by the non-adaptive combiners. Some combination classifiers that have been used in Indian scripts are ANN and HMM [61], K-Means and SVM [54], MLP and SVM [54], MLP and minimum edit, SVM and ANN, fuzzy neural network [36], NN, fuzzy logic and genetic algorithm [52]. Pawan Kumar [30] used five different classifiers (two HMM and three NN based) to obtain a better accuracy.

3.6 Post Processing

It is a step that has been performed after the scanning and recognition of the captured image of handwritten characters. While processing this step, there can be two types of errors.

3.6.1 Real Word Errors

This type of error is meaningful, but contains a non-required word for the particular sentence. Those type of errors are the complex ones and can disturb the overall syntax and semantic of the sentence. For the detection, human involvement is necessarily required because automatic correction of syntax and semantics for this type of errors can be a critical task for a program or machine. For example, when -what are you doing is recognized by the handwritten characters recognition system as -what are your doing , here-your is considered as a real word error because -your is correct but grammatically incorrect for this sentence [41].

3.6.2 Non-word Errors

This type of errors doesn't have any particular meaning in the sentence. There are many algorithms that have been proposed by the researchers for the detection and to provide suggestion words for the correction of the recognized contents. Algorithms like spell-checker are commonly used for this type of errors. For example, when -Are you happy is fed to the recognition by the handwritten character recognition system as, -Are you appy, here in this sentence -appy is the non-word error [36],[37].

4. Applications

- Automative Form entry systems for various paper works like: office, medical, legal authorities, Education, Finance and Government agencies.
- Automatic recognition of number plates.
- Automative postal address recognition from letters.
- To make texual documents of the printed documents.
- For technological growth of blind and impaired users.
- Cost, time and manpower reduction.

Conclusion

In this paper, the authors have presented a detailed discussion about the handwritten character recognition, which contains various concepts that are being applied in this field. The rate of handwritten characters recognition accuracy basically depends on nature and shape of the handwritings of individuals. In this field, the current research is concentrated on recognition of cursive handwritings, multilingual mixed type handwritten characters and for the child written characters, for which high supervision is needed. From various research papers, the authors have found that the selection of appropriate feature extraction and classification methodologies personate a vital role for the performance and accuracy of the handwritten characters recognition.

References

[1]. A. Kundu, and Y. He, (1991). "On optimal order in modeling sequence of letters in words of common language as a Markov chain". *Pattern Recognit.*, Vol.24, No.7, pp.603-608.

[2]. Anoop M. Namboodiri, and Anil K. Jain, (2004). "Online Handwritten Script Recognition". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.26, No.1.

[3]. Anil K. Jain, Robert P.W. Duin, and Jianchang Mao, (2000). "Statistical Pattern Recognition: A Review". *IEEE*

Transactions on Pattern Analysis and Machine Intelligence, Vol.22, No.1, pp.4-37.

[4]. Arjun Singh, and Kansham Angphun Maring, (2015). "Handwritten Devanagari Character Recognition using SVM and ANN". *International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 4, No. 8, pp. 123-128.

[5]. Anshul Gupta, Manisha Srivastava, and Chitralekha Mahanta, (2011). "Offline Handwritten Character Recognition Using Neural Network". *IEEE (ICCAIE)*, pp. 102-107.

[6]. B. Hari Kumar, Y.E. Vasanth Kumar and Mohammad Shaffi, (2014). "Optical character recognition using Split Profile Algorithm". *International Journal of Advanced Engineering and Global Technology*, Vol.2, No.12.

[7]. Brijmohan Singh, Ankush Mittal, M.A. Ansari and Debashis Ghosh, (2011). "Handwritten Devanagari Word Recognition: A Curvelet Transform Based Approach". *IJCSE*, Vol.3, No. 4, pp. 1658- 1665.

[8]. Chanchal Bansal, and Arif Khan, (2014). "Handwritten Numeral Recognition using SVM and Chain Code". *IJARET*, Vol.2, No.7.

[9]. C.Y. Suen, M. Berthod, and S. Mori, (1980). "Automatic recognition of handprinted characters—The state of the art". *Proc. IEEE*, Vol.68, pp.469-487.

[10]. C. V. Jawahar, M. N. S. S. K. Pavan Kumar and S. S. Ravi Kiran, (2003). "A Bilingual OCR for Hindi-Telugu Documents and its Applications". *Seventh International Conference on Document Analysis and Recognition (ICDAR)*.

[11]. Deepa Berchmans, and S.S. Kumar, (2014). "Optical Character Recognition: An Overview and an Insight". *IEEE, International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT),* pp. 1361-1365.

[12]. Divakar Yadav, Sonia Sánchez-Cuadrado and Jorge Morato, (2013). "Optical Character Recognition for Hindi Language Using a Neural-network Approach". *Journal of Information Processing Systems*, Vol.9, No.1, pp.117-140.

[13]. Gaurav Y. Tawde, (2014). "Optical Character Recognition for Isolated Offline Handwritten Devanagari Numerals Using Wavelets". *Journal of Engineering Research and Applications*, Vol.4, No.2.

[14]. H. Cheng, W. H. Hsu, and M. C. Kuo, (1993). "Recognition of handprinted Chinese characters via stroke relaxation". *Pattern Recognit.*, Vol.26, No.4, pp.579-593.

[15]. I.S. Oh and C. Y. Suen, (1998). "Distance features for neural network-based recognition of handwritten characters". *Int. J. Document Anal.Recognit.*, Vol.1, No.2, pp.73-88.

[16]. J. Rocha and T. Pavlidis, (1994). "A shape analysis model". *IEEE Trans. Pattern Anal. Machine Intell.*, Vol.16, pp.394-404.

[17]. J. Kitler, (1986). "Feature selection and extraction". *In Handbook of Pattern Recognition and Image Processing.* T. Young and K. Fu, Eds, NewYork: Academic, pp.59-83.

[18]. J. Hu, and T. Pavlidis, (1996). "A Hierarchical Approach to Efficient Curvilinear Object Searching". *Computer Vision and Image Understanding,* Vol.63, No.2, pp.208-220.

[19]. K. Y. Rajput and Sangeeta Mishra, "Recognition and Editing of Devnagari Handwriting Using Neural Network". *SPIT-IEEE Colloquium and Intl. Conference*, Mumbai, India.

[20]. K. Chung and J. Yoon, (1997). "Performance comparison of several feature selection methods based on node pruning in handwritten character recognition". *In Proc. 4th Int. Conf. Document Anal. Recognit.,* Ulm, Germany, pp.11-15.

[21]. K. Wang, Y.Y. Tang, and C.Y. Suen, (1988). "Multi-layer projections for the classification of similar Chinese characters". In *Proc. 9th Int. Conf. Pattern Recognit.*, pp.842-844.

[22]. K.T. Miura, R. Sato, and S. Mori, (1997). "A method of extracting curvature features and its application to handwritten character recognition". In *Proc. 4th Int. Conf. Document Anal. Recognit.*, Ulm, Germany, pp.450-454.

[23]. M.D. Garris, et al., (1994). "NIST form based handprinted recognition system". Nat. Inst. Standard

Technol., Gaithersburg, MD, *Tech. Rep. NISTIR 546 920 899.*

[24]. M. Egmont-Petersen, and D. De Ridder, H. Handels, (2002). "Image Processing with Neural Networks: A Review". *Pattern Recognition*, Vol.35, pp.2279-2301.

[25]. M. Hamanaka, K. Yamada, and J. Tsukumo, (1993). "On-Line Japanese Character Recognition Experiments by an Off-Line Method Based on Normalization-cooperated Feature Extraction". In *Proc. 3rd ICDAR*, pp.204-207.

[26]. M.K. Brown and S. Ganapathy, (1983). "Preprocessing techniques for cursive script word recognition". *Pattern Recognit.*, Vol.16, No.5, pp.447-458.

[27]. Munish Kumar, M.K. Jindal and R. K. Sharma, (2011). "k -Nearest Neighbor Based Offline Handwritten Gurmukhi Character Recognition". *IEEE ICIIP*.

[28]. Mohamed Cheriet, Nawwaf Kharma, and Cheng-Lin Liu, Ching Y. Suen, (2007). *Character Recognition Systems. A Guide for Students and Practioners*. John Wiley & Sons, Inc.

[29]. Malakar, Samir, et al., (2012). "Text line extraction from handwritten document pages using spiral run length smearing algorithm". *Communications, Devices and Intelligent Systems (CODIS), International Conference on, IEEE.*

[30]. M N S S K Pavan Kumar, S S Ravikiran, Abhishek Nayani, C V Jawahar and P J Narayanan, (2003). "Tools for Developing OCRs for Indian Scripts".*Computer Vision and Pattern Recognition Workshop, 2003, CVPRW'03, Conf. on, IEEE,* Vol. 3, pp. 33.

[31]. Monica Patel, and Shital P. Thakkar, (2015). "Handwritten Character Recognition in English: A Survey". *International Journal of Advanced Research in Computer and Communication Engineering*, Vol.4, No.2.

[32]. Mohammad Abu Obaida, Md. Jakir Hossain, Momotaz Begum and Md. Shahin Alam, (2011). "Multilingual OCR (MOCR): An Approach to Classify Words to Languages". *International Journal of Computer Applications*, Vol.32, No.1.

[33]. Nabin. Sharma, U. Pal, F. Kimura, and S. Pal, (2006). "Recognition of Off-Line Handwritten Devnagari Characters Using Quadratic Classifier". Springer-Verlag, Berlin Heidelberg ICVGIP, LNCS 4338, pp.805-816.

[34]. Nafiz Arica, and Fatos T. Yarman-Vural, (2001). "An Overview of Character Recognition Focused on Off-Line Handwriting". *IEEE Transactions on Systems*, Man, and *Cybernetics—Part C: Applications and Reviews,* Vol.31, No.2.

[35]. N. Otsu, (1979). "A threshold selection method from gray-level histograms". *IEEE Transactions on Systems, Man, and Cybernetics SMC*, Vol.9(1), pp.62-66.

[36]. N. Arica and F.T. Yarman-Vural, (2000). "One dimensional representation of two dimensional information for HMM based handwritten recognition". *Pattern Recognit. Lett.*, Vol.21, No.6-7, pp.583-592.

[37]. Nafiz Arica, (1998). "An Offline Character Recognition System for Free Style Handwriting". (Doctorial Dissertation), The Middle East Technical University.

[38]. Ohhira T., Pecharanin N., Taguchi A., Iijima, N., Akima, Y., and Sone, M., (1995). "Chinese character recognition by the auto recognition system". *IEEE International Conference on Neural Networks*, Vol.5, pp.2222-2225.

[39]. Paul W. Handel, (1993). *Statistical Machine.* U.S. Patent 1915993, Retrieved From http://www.google.com/patents?vid=USPAT1915993.

[40]. P.S. Deshpande, Latesh Malik, and Sandhya Arora, (2007). "Recognition of Handwritten Devanagari Characters with Percentage Component Regular Expression Matching and Classification Tree". *TENCON 2007-2007, IEEE Region 10 Conference*, pp.1-4.

[41]. P.D. Gader and M. A. Khabou, (1996). "Automatic feature generation for handwritten digit recognition". *IEEE Trans. Pattern Anal. Machine Intell.*, Vol.18, pp.1256-1262.

[42]. R.J. Ramteke, and S.C. Mehrotra, (2008). "Recognition of Handwritten Devnagari Numerals". *International Journal of Computer Processing of Oriental Languages*.

[43]. Rekha Singh, (2014). "Character Recognition for Bi-Lingual Mixed-type Characters using Artificial Neural

Network". *IJRET: International Journal of Research in Engineering and Technology*, Vol.3, Special Issue.10.

[44]. S.S. Wang, P.C. Chen, and W.G. Lin, (1994). "Invariant pattern recognition by moment Fourier descriptor". *Pattern Recognit.*, Vol.27, pp.1735-1742.

[45]. S. Madhvanath and V. Govindaraju, (1999). "Reference lines for holistic recognition of handwritten words". *Pattern Recognit.*, Vol.32, No.12, pp.2021-2028.

[46]. S. Madhvanath, G. Kim, and V. Govindaraju, (1999). "Chaincode contour processing for handwritten word recognition". *IEEE Pattern. Anal. Machine Intell.*, Vol.21, pp.928-932.

[47]. S. Madhvanath, E. Kleinberg, V. Govindaraju, and S. N. Srihari, (1997). "The HOVER system for rapid holistic verification of off-line handwritten phrases". In *Proc. 4th Int. Conf. Document Anal. Recognit.*, Ulm, Germany, pp.855-890.

[48]. S.W. Lee and Y.J. Kim, (1995). "Direct extraction of topographic features for gray-scale character recognition". *IEEE Trans. Pattern Anal. Machine Intell.*, Vol.17, pp.724-729.

[49]. Sandhya Arora, et al., (2010). "Performance Comparison of SVM and ANN for Handwritten Devanagari Character Recognition". *IJCSI International Journal of Computer Science Issues*, Vol.7, No.3.

[50]. Sandhya Arora, Debotosh Bhattacharjee, Mita Nasipuri, D. K. Basu and M. Kundu, (n.d.). " Recognition of Non-Compound Handwritten Devanagari Characters using a Combination of MLP and Minimum Edit Distance". *International Journal of Computer Science and Security (IJCSS)*, Vol.4, No.1, pp.107-120.

[51]. Shailedra Kumar Shrivastava, and Sanjay S. Gharde, (2010). "Support Vector Machine for Handwritten Devanagari Numeral Recognition". *International Journal of Computer Applications*, Vol.7, No.11.

[52]. Sung-Bae Cho, (2002). *Fusion of Neural Networks with Fuzzy Logic and Genetic Algorithm*. IOS Press, pp.363-372.

[53]. Sanjay Chandra Arya, Rajesh Shyam Singh and Hardwari Lal Mandoria, (2015). "Image Denoising in Hand Written Document for Degraded Documents using Wiener Filter Algorithm". *International Journal for Research in Emerging Science and Technology*, Vol.2, No.7.

[54]. Satish Kumar, (2009). "Performance and Comparison of Features on Devanagari Hand-printed Dataset," *Inlt. Journal of Recent Trends in Engineering*, Vol.1, No.2.

[55]. Sandhya Arora, Debotosh Bhattacharjee, Mita Nasipuri, M. Kundu and D. K. Basu and L. Malik, (2010). "Handwritten Devanagari Numeral Recognition using SVM & ANN". *IJCSE40*, Vol.1, No.2.

[56]. I.K. Sethi and B. Chatterjee, (1976). "Machine Recognition of hand printed Devanagri numerals". *J. Inst. Electron Telecommun, Engg,* pp. 532-535.

[57]. Saurabh Farkya, Govinda Surampudi, and Ashwin Kothari, (2015). "Hindi Speech Synthesis by concatenation of recognized Hand written Devanagri Script using Support Vector Machines Classifier". *IEEE ICCSP Conference*.

[58]. R.M.K. Sinha, and H.N. Mahabala, (1979). "Machine Recognition of Devanagri Script. *IEEE Trans. System. Man Cybern*, pp. 435-441.

[59]. Tapan K Bhowmik, Swapan K Parui Utpal Roy, (2008). "Discriminative HMM Training with GA for Handwritten Word Recognition". *IEEE*.

[60]. Teena Mittal, and Rajendra Kumar Sharma, (2015). "Multiclass SVM based Spoken Hindi Numerals Recognition". *International Arab Journal of Information Technology*, Vol.12, No.6A.

[61]. Umapada Pal, Sukalpa Chanda Tetsushi, Wakabayashi, and Fumitaka Kimura, (n.d.). "Accuracy Improvement of Devnagari Character Recognition Combining SVM and MQDF".

[62]. U. Pal, T. Wakabayashi, and F. Kimura, (2009). "Comparative Study of Devnagari Handwritten Character Recognition using Different Feature and Classifiers". *10th Intl. Conf. on Document Analysis and Recognition,* pp.1111-1115.

[63]. Vikas J Dongre, and Vijay H Mankar, (2010). "A

Review of Research on Devnagari Character Recognition". *International Journal of Computer Applications,* Vol.12, No.2, pp.0975- 8887.

[64]. Vijay Laxmi Sahu, and Babita Kubde, (2013). "Offline Handwritten Character Recognition Techniques using Neural Network: A Review". *International Journal of Science and Research (IJSR)*, Vol.2, Issue.1.

[65]. W. Lu, Y. Ren, and C. Y. Suen, (1991). "Hierarchical attributed graph representation and recognition of handwritten chinese characters". *Pattern Recognit.*, Vol.24, No.7, pp.617-632.

[66]. W.Y. Kim and P. Yuan, (1994). "A practical pattern recognition system for translation, scale and rotation invariance". In *Proc. Int. Conf. Comput. Vis.Pattern Recognit.*, Seattle, WA.

[67]. X. Li, W. Oh, J. Hong, and W. Gao, (1997). "Recognizing components of handwritten characters by attributed relational graphs with stable features". In *Proc. 4th Int. Conf. Document Anal. Recognit.*, pp.616-620.

[68]. X. Zhu, Y. Shi, and S. Wang, (1999). "A new algorithm of connected character image based on Fourier transform". In *Proc. 5th Int. Conf. DocumentAnal.*

Recognit., Bangalore, India, pp.788-791.

[69]. Y. Tao and Y. Y. Tang, (1999). "The feature extraction of Chinese character based on contour information". In *Proc. 5th Int. Conf. Document Anal.Recognit.*, Bangalore, India, pp.637-640.

[70]. Y. C. Chim, A. A. Kassim, and Y. Ibrahim, (1999). "Character recognition using statistical moments". *Image Vis. Comput.*, Vol.17, pp.299-307.

[71]. D. Trier, A.K. Jain, and T. Taxt, (1996). "Feature extraction method for character recognition—A survey". *Pattern Recognit.*, Vol.29, No.4, pp.641-662.

[72]. (1995). "Multiresolutional recognition of handwritten numerals with wavelet transform and multilayer cluster neural network". In *Proc. 3rd Int. Conf. Document Anal. Recognit.*, Montreal, QC, Canada, pp.1010-1014.

[73]. D. Guillevic and C.Y. Suen,(1998). "Recognition of legal amounts on bank cheques". *Pattern Anal. Applicat.*, Vol.1, No.1, pp.28- 41.

[74]. (1998). "HMM–KNN word recognition engine for bank cheque processing". In *Proc. 14th Int. Conf. Pattern Recognit.*, Brisbane, Australia, pp.1526-1529.

ABOUT THE AUTHORS

Ujwal Singh Vohra presently pursuing her M.Tech in the Department of Information Technology from G.B.P.U.A.T. Pantnagar, India. She received her B.Tech degree in Information Technology from Govind Ballabh Pant University of Agriculture and Technology Pantnagar, India. Her research interest includes Image Processing, Animation and Search Engine Optimization.

Shri Prakash Dwivedi is currently working as an Assistant Professor in the Department of Information Technology at College of Technology GB Pant University of Agriculture & Technology, Pantnagar, India. His areas of Interest are Algorithms and Theory of Computing.

Dr.Hardwari Lal Mandoria is currently working as a Professor & Head in the Department of Information Technology at College of Technology GB Pant University of Agriculture & Technology, Pantnagar, India. His areas of Interest are Computer Networks, Network Security Wireless Communication, Mobile Computing, Information Security and Information Communication Technology.