

# Scene Text Recognition using Artificial Neural Network: A Survey

Sunil Kumar  
School of Computer  
Applications,  
IFTM University, Moradabad,  
India

Krishan Kumar, PhD  
Department of Computer  
Science,  
Gurukul Kangri University,  
Haridwar, India

Rahul Kumar Mishra, PhD  
School of Computer  
Applications,  
IFTM University,  
Moradabad, India

## ABSTRACT

Nowadays, scene text recognition has become an important emerging area of research in the field of image processing. In image processing, character recognition boosts the complexity in the area of Artificial Intelligence. Character recognition is not easy for computer programs in comparison to humans. In the broad spectrum of things, it may consider that recognizing patterns is the only thing which humans can do well and computers cannot. There are many reasons including various sources of variability, hypothesis and absence of hard-and-fast rules that define the appearance of a visual character. Hence; there is an unavoidable requirement for heuristic deduction of rules from different samples. This review highlights the superiority of artificial neural networks, a popular area of Artificial Intelligence, over various other available methods like fuzzy logic and genetic algorithm. In this paper, two methods are listed for character recognition – offline and online. The “Offline” methods include Feature Extraction, Clustering, and Pattern Matching. Artificial neural networks use the static image properties. The online methods are divided into two methods, k-NN classifier and direction based algorithm. Thus, the scale of techniques available for scene text recognition deserves an admiration. This review gives a detail survey of use of artificial neural network in scene text recognition.

## Keywords

Character Recognition, Scene text recognition, Text extraction, Feature extraction, Artificial Neural Network.

## 1. INTRODUCTION

Image recognition is a very important area of research. Many methods are available for different types of image recognition such as 2D and 3D [1]. But due to the increasing demand of security and new technology enhancement it is great need of text recognition. And hence the scene text recognition is also an interesting and important area of research.

It is well known that text recognition has become an important issue of daily life. It can be seen in different places for the purpose of security. Another area of scene text recognition is the extension of text recognition. It can be recognized by many methods used so far in the image processing research. But soft computing techniques like artificial neural networks, genetic algorithm, and fuzzy logic are the widely and most commonly used techniques.

In the rich milieu of real world problems which are not interpreted or too complicated to interpret mathematically, an integration of methodologies that were designed to model and enable solutions came into light that is soft computing technique. Soft computing in other words is a combination of synergistically working methodologies which provides, in any form, flexibility of handling real-life ambiguous situations and malleable information processing potential for handling [2].

The computation methods give the approximate solution to an imprecisely or precisely formulated problem [3]. Soft computing is different from hard (conventional) computing as soft computing has tolerant of imprecision, partial truth, uncertainty and approximation. In effect, the human mind is the backbone for soft computing. Soft computing is a technique which gives the solution of very hard machine learning problems.

With fast escalation of multimedia documents and demand for information indexing and retrieval, much venture has been done on scene text from images and videos [4]. The main guerdon of such an attempt was not only human studying but also the probability of systematic application in which handwritten and/or printed character available on document has to be transformed into machine learning design [5, 6]. Automatic character recognition of printed and handwritten information in various commercial departments such as libraries, post offices, banks, and publishing houses on documents like cheques, envelopes, other manuscripts and application forms of different filed has a variety of applications [7]. Now a days, Optical Character Recognition (OCR) and artificial neural network (ANN) is used for character verification and classifier respectively.

Optical character recognition (OCR) is the area of research in artificial intelligence, pattern recognition, image processing and machine learning [8]. The conversion of handwritten or printed document file into editable text format takes place via a mechanism, optical character recognition. Now, an argument about the area of video text where two types of text in the video: Scene text and Caption text. Text on container, signs, CD cover, banners, sign board and text on vehicle are examples of scene text or graphics text. On the other hand, subtitles in news video, the scores in sports videos and date and time in the video are the some examples of caption text or artificial text in which text is artificially overlaid on the video/image. It may be defined as superimposed text. However, the extraction of text from image or video is a challenging task due to variations of text in differences in font, style, size, alignment, orientation, unknown layout and complex background.

### A. Scene Text Recognition: An insight

Text detection and recognition is still a demanding and unsolved computer vision problem in natural scene images. Scene text has image noise, image blur, complex background, partially occluded text, varying illumination and variations in font-styles, and size as represented by Figure 1.

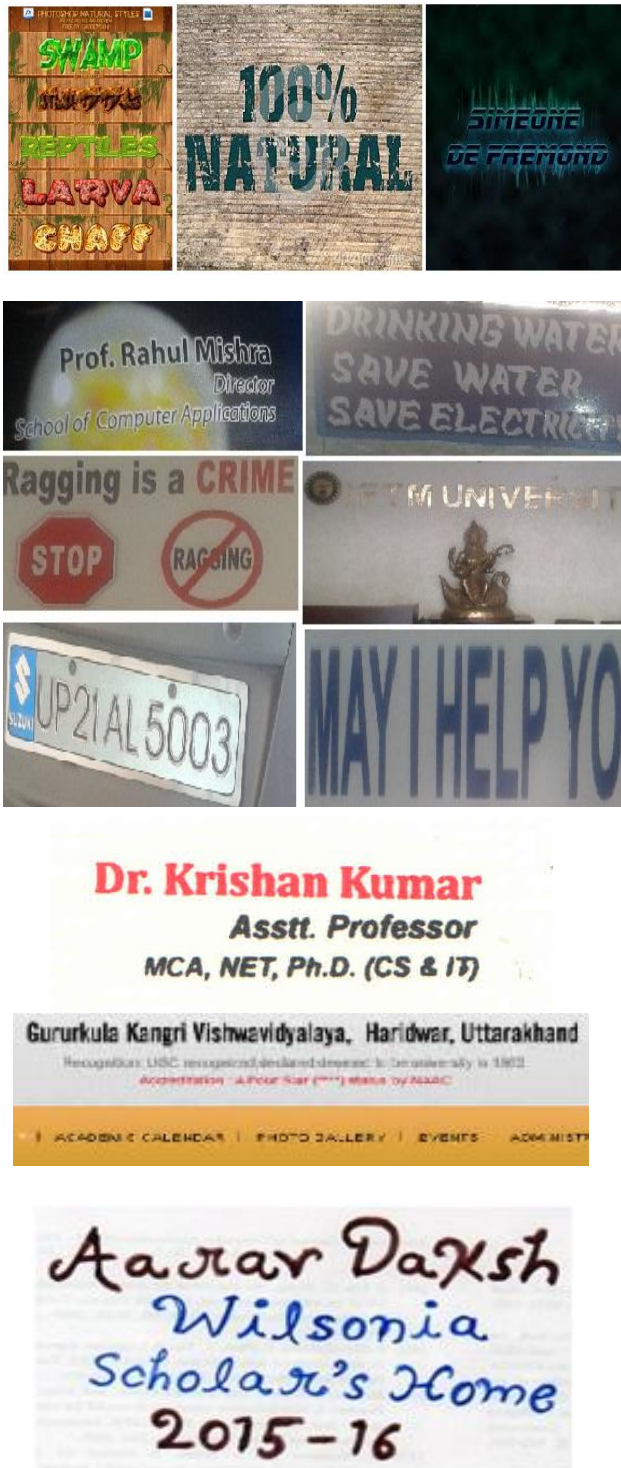


Fig1. Scene images containing text

The first motive of text localization is to locate individual characters, words or lines of text. Then, if one knows the location of regions of text in the image, should further look into identification of the actual words and lines of text in located regions. An illustration of the end-to-end text location and recognition task is represented by Figure. 2.

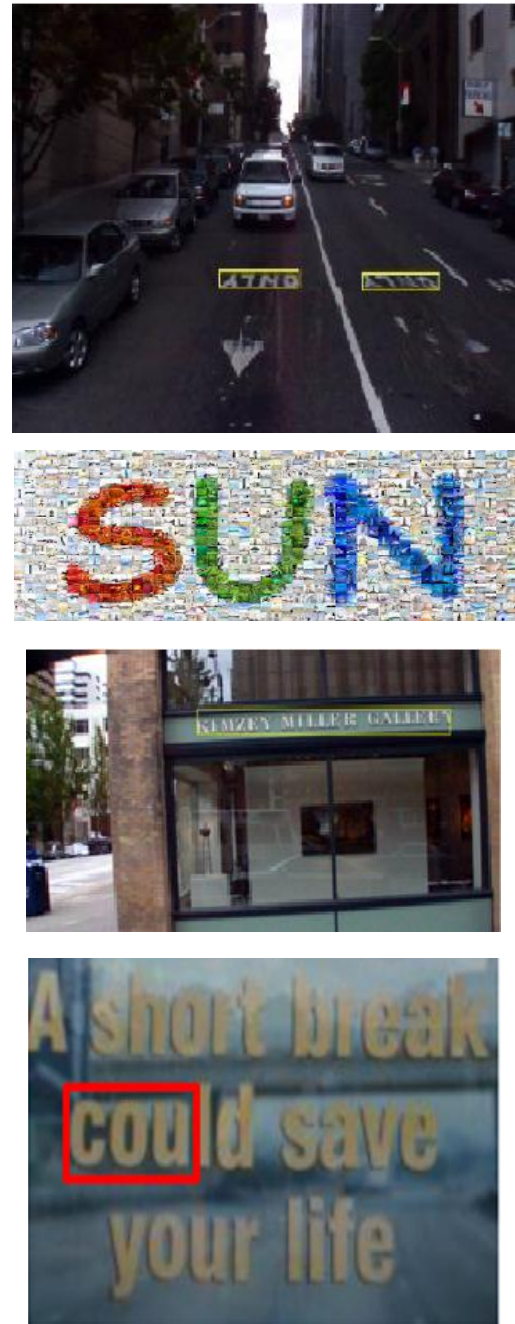


Fig. 2 Illustration of the end-to-end text location and recognition

Now, various existing algorithms are able to achieve extremely high resolution performance on specialized tasks such as digit recognition in stiff awkward settings. For illustration, the system is biologically plausible deep artificial neural network architectures are able to achieve matching of human performance on tasks such as the recognition of handwritten digits or traffic signs [9], [10], [11]. Similarly, the system applied recently developed methods in machine learning concretely, substantial procedure for learning the features naturally from unlabeled data and concluded that humans can construct highly effective classifiers for both text recognition and detection to be used in a more accuracy end-to-end system [12]. Although there is a list of advances in the area of text recognition, however, the more general task of detecting and recognizing text in complicated scenes still remains an open challenge.

## 2. RELATED WORK

*Chuai Yi et al.* proposed a novel algorithm, on the grounds of stroke components and descriptive Gabor filters and intention was to detect text regions in scene images. Author approved that the statistical analysis on the stroke factor of text from training sets to acquire SGWs, which are used as the entire description of principle stroke components. Rayleigh model is used for traditional typical attributes distributions for every SGW to describe suitability statistics on positive training samples and the Gaussian model to describe suitability statistics on negative training samples [13].

*Xu-Cheng Yin et al.* proposed the strategy of minimizing regularized variations to extract Maximally Stable Extremal Regions (MSERs) where distance weights and threshold of the clustering algorithm are classified for text candidates similar to self-training distance metric learning algorithms that enables the probabilities of text candidates related to non-text candidates as a character classifier, and finally texts are marked into a text classifier [14].

*Parshuram M. Kamble et al.* proposed representation of RHOG (Rectangle Histogram Oriented Gradient) as the background for extraction of characteristics by the support of SVM (Support Vector Machines) and FFANN (feed-forward Artificial Neural Network) classification technology for measurement of arithmetic operations per image pixel which makes them acceptable for real-time applications and also to demonstrate the high performance [15].

*Yi-Feng Pan et al.* designed a local binarization approach to create text confidence map with the help of text region detector. The labeled components as “text” or “non-text” are presented in Conditional Random Field (CRF) model related to unary component properties with neighboring components. Last, text components are grouped into text lines with an energy minimization approach. This method was proposed to promise the performance for comparing with the existing methods on ICDAR 2003 competition dataset [16].

*A. Thilagavathy et al.* put forward a distinctive hybrid approach consisting of two most well-similar text extraction technologies. This approach involves two types of methods, first is region based method and other is connected component (CC) based method. In this research, first of all, the video is divided into frames and key frames received. Candidate region is developed to perform binarization and text region indicator (TRI) is being augmented to measure the text prevailing confidence. Here ANN and OCR (Optical Character Recognition) are used as the classifier and character verifier respectively. Text is gathered by generating the minimum spanning tree with the implementation of bounding box distance [17].

*Xiangrong Chen et al.*, proposed AdaBoost algorithm used for detecting text in unconstrained city scenes. City scenes are very common in movies and normal videography and photography. The key element was the choice of the feature set which was selected to have features with low entropy of the positive training examples [18].

*Marc'Aurelio Ranzato et al.*, illustrate a novel method which uses a linear encoder and decoder introduced by a sparsifying non-linearity that transforms a code vector to a quasi-binary sparse code. Disposal of an input is followed by the main task of optimal code to diminish the distance between the input patch and output of the decoder while being as equivalent as possible to the encoder output [19].

*B. Epshtein et al.* present a novel image operator that explores to detect the value of stroke width for each image pixel, and demonstrate the text detection in the images. The commend operator is local and data dependent, which define it fast and sturdy adequate to eliminate the demand for multi-scale

scanning or computation windows. A vast testing demonstrates that the proposed scheme defeats the latest published algorithms and simplicity empowers the algorithm to detect texts in multiple languages and fonts [20].

*R. Manmatha et al.* suggested an algorithm for smooth word segmentation of traditional handwritten documents. Numerous organizations, libraries and other museums, comprise a huge collection of handwritten conventional documents. Contemporary segmentation techniques such as gap metrics algorithm have been mostly succeed and assessed on extremely pressurized documents like postal addresses and bank checks. An insignificant work has been done and reported on completely handwritten pages and this work has usually includes screening on clean artificial documents specially designed for the enthusiasm of research. Traditional manuscript images are much more exciting due to implication of a great deal of noise. A scale space algorithm for automatically separated classical handwritten documents into words is illustrated. In the first instance, the page is cleaned to eliminate margins followed by application of gray-level projection profile algorithm in search of lines in images. All line images are then sieved by an anisotropic Laplacian at diverse scales. The strategy furnishes blobs which correlate with fragments of characters at small scales and to words at higher scales. Censorious to the algorithm is choice of scale that is, to find the most appreciative scale at which blobs consistent with words. This is completed by detecting the extreme over scale of the zone or extent of the blobs. The use of three different strategies projects this extreme scale. The recovery of blobs at the most convenient scale is followed by bounding with a rectangular box in order to reveal the words. The elimination of boxes of unfamiliar size which are unlikely to correspond to words can be done by an after-processing filtering step. The method is accomplished on a lot of divergent data sets and it has been proven that, on 100 sampled documents from the George Washington corpus of handwritten document images, 17 percent of total error rate is resolved. The technique outstandingly performs a recent gap metrics word-segmentation algorithm on this phase [21].

*M. K. Jindal et al.* suggested an explanation for segmentation of horizontally overlapping lines and interpret the difficulty of eight widely used common printed Indian scripts. In this explanation, entire document is segmented into strips followed by utilization of presented algorithm for segregating horizontally overlapping lines and combining small strips to their original lines [22].

*Abderrazak Zahour et al.* brought to light an alternative method for text-line segmentation stand on Block Covering which cracks the complication of multi-touching and overlapping constituents. Block Covering is the core of a system which practices a set of foregone Arabic documents from authentic evidences. The system is depicted for the motive of segmenting text-lines unlikely if they are touching at various stages and overlapping. The Block Covering technology is capitalized on three steps. Step one includes document classification by Block Counting, a novel way of fractal analysis. Step second consists of Block classification by a statistical probe of block heights. Finally step three is a neighboring examination for generating text lines. A fuzzy C-means procedure in association with the Block Counting fractal analysis, is used to achieve document images in order to categorize them on the grounds of their complexity: closely (tightly) packed documents (TPD) or widely (broadly) packed documents (WPD). An optimal Block Covering is performed on TSD documents consisting of overlapping and multi touching and mixed text lines. The titanic blocks produced by the covering are further segmented by depending on the

statistical examination of block heights. The ultimate labeling into text-lines is relying on a block nearby analysis [23].

*G. Louloudis et al.* has highlighted a text line detection approach for handwritten documents. The proposed technique is stand on a method that comprises of three discrete stages. The initial step consists of pre-processing of input image and connected component removal, separation of the connected component domain into standard character height approximation and three spatial sub-domains. Secondly, author applied a block-based Hough transform to detect the probable text lines while on the other hand, third step includes correction of probable splitting, to recognize text lines that the starting and second steps could not expose and, finally, to discontinue vertically associated characters and allocates them to text lines [24].

*Nikolaos Stamatoopoulos et al.* has suggested a combining approach of various segmentation methods. It is done for the sake of squeeze the segmentation outcome of complementary approaches and particular attributes of the starting image so as to create boosted segmentation consequences. This association method is consists of five steps: Average attribute extraction, to detect accurately segmented areas, division of small areas into groups, generate accurately segmented areas from every group and finally procedure of the new segmentation outcome [25].

*Yi Li et al.* has suggested a method stand on density approximation and a ultra-modern image segmentation approach, the level set technique. A probabilistic map is approximate from an input document image where every element determines the expectation of the underlying pixel 12 related to a text line. In addition, level set approach is evolved to regulate the boundary of adjacent text lines by developing an elementary estimate [26].

*A. Nicolaou et al.* proposed a method to slice handwritten images into text lines by shredding their surface with local minima tracer. It is suppose that there exists a pave from one side to other of the image that traverses wholly single text line. first of all Image is blurred and then uses tracers to follow the white-most and black-most paths from both left to right and right to left direction in order to shred the image into text line areas [27].

*Alex Graves et al.* proposed an alternative approach based on a novel type of recurrent neural network, specially designed for sequence labeling tasks where the data is hard to segment and contains long-range bidirectional interdependencies. In experiments on two large unconstrained handwriting databases, this approach achieved word recognition accuracies of 79.7 percent on online data and 74.1 percent on offline data, appreciably outperforming a state-of-the-art HMM-based system [28].

*F. Ahmed et al.* has further reported the Niblack's approach for segmentation. It is the type of local thresholding segmentation techniques. The output of Niblack's approach is of considerable interest and has most satisfactory consequence in the class of all thresholding techniques for segmenting text documents. In this research, same process is utilized on images, keeping one of all variables constant and the variable is weight  $k$  of Niblack's approach but changing the other window size variable from one image to other. The resultant image is more segmented nonetheless the background of image is noisy. Augmentations in the outcome images are manifested by utilizing the closing and opening morphological behavior. Closing and opening are categories of two underlying morphological activities erosion and dilation. Erosion contracts matter in a binary image on the other hand dilation condenses objects in a binary image by the addition of pixels to the frontier of the objects [29].

*Fei Yin et al.* highlighted a text line segmentation algorithm on the grounds of minimal spanning tree (MST) congregation with distance metric learning method. The connected components (CCs) of document or scene images are categorized into a tree like structure, from distance metric text lines are filtered by dynamically cutting the edges utilizing a contemporary hyper volume depletion criterion and a straightness measurement. The suggested algorithm covers numerous documents with curved and multi-skewed text lines [30].

*Xiaojun Du et al.* reported a novel text line segmentation procedure stand on the Mumford-Shah model. The algorithm does not correlate with script, implements piecewise persistent approximation of the MS model in order to segment handwritten and/or scene text images. Additionally, author further used morphing to eliminate overlaps between adjacent text lines and to associate broken text lines [31].

*Azizah Suliman et al.* provided detailed descriptions of the algorithms used in the preprocessing and feature extraction phases of an offline handwritten character. In classifying handwritten characters, the stages prior to the classification phase play a role as major as the classification itself. The whole research presents a hybrid approach of HMM and Fuzzy Logic in the field of handwritten character recognition. Fuzzy Logic is used in the classification phase while HMM is used in the process of extracting features for the preparation of linguistic variables of the fuzzy rules. However, only the preprocessing stages as employed by the research are described here. The pre-processing phase starts from reading in the input file, the process of binarization, reference line estimation and thinning of the character image for further use in the next stage of the feature extraction and recognition process. Each of the pre-processing stages and the chain coding process will be described in detail giving improvised algorithms, and examples of the processes on existing samples from the database shown [32].

*Vassilis Papavassiliou et al.* reported two techniques to express words and text lines from long handwritten document. The line segmentation approach is stand on locating the optimal accession of text and gap zones within vertical zones by applying the Viterbi approach. With the help of text-line separation drawing approach the connected components are assigned to text lines [33].

*Alireza Alaei et al.* proposed a painting scheme to accomplish the objective of line segmentation of unconstrained handwritten text. The newly approach has been devised by studying the cursive Persian text scripts broadly. The line segmentation algorithm approach is relevant to handwritten text in any script. Each strip in every row is painted with the average gray intensity value of whole pixels intensity available in that particular row-strip. Later, the painted pipes are converted into two-tone painting and then smoothed. The white/black places in each pipe of the smoothed image are analyzed to get a short line of separation called as Piece-wise Potential Separating Line (PPSL), between two consecutive black spaces. The PPSLs are joined to produce the segmentation of text lines [34].

*B M Singh et al.* Fast approaches are necessarily required for effective image processing systems for conducting huge data for calculations. With the help of Graphics Processing Unit (GPU) speed up the processing and parallel implementation of an algorithm can be done. GPU is programmability; general purpose computation hardware and low cost make it productive. Binarization method is widely used in the image recognition and analysis. Author purpose the accuracy and performance parameters of GPUs on popular, global binarization Otsu's approach for OCR. The main intention of

this purpose work is to create binarization faster for recognition of a huge number of document images on GPU [35].

*Dharmapryia C. Bandara et al.* suggested an algorithm that assures the segmentation of discrete handwritten character segments into meaningful skeletons, ignoring the difficulties of above-segmentation and below-segmentation. Fuzzy handwritten character recognition systems have the attributes of effectiveness, online adaptability and flexibility. A crucial necessity of such a fuzzy system for either offline or online handwritten character recognition is the breaking up meaningless individual characters into skeleton of meaningful segments. Now, these segments may be applied for the calculation of fuzzy hallmarks. Further these attributes can be applied for the recognition of individual character. This paper expresses a developed segmentation algorithm for segmentation of individual offline handwritten characters [36].

*Pezeshk et al.* proposed a system which was based upon automatic feature extraction and text recognition from scanned topographic maps. They showed that the recovered text was recognized using this multifont segmentation free optical character recognition (OCR). The custom OCR has the capability to achieve a 94% recognition rate for the extracted text, thereby outperforming a commercial OCR used as a benchmark [37].

*Tanay Kumar Roy et al.* proposed that skewness of images can be corrected for any degree of rotation by using an OJ approach. In continuation, the paper deals with Stroke-Whitespace distance approach for the sake of 180° rotation of images. As the technology develops, requirement of Optical Character Recognition (OCR) for both documents (printed and handwritten) of any language is unavoidable. In the procedure of improving an OCR for Hindi, Marathi and Bengali etc. languages, that are top 15 amongst all, the widely spoken languages of the world, the intention of skewness correction has been carried out in the field and still remains a challenging task among researchers. They acquire the challenge and presented a algorithmic approach rely on stroke-whitespace that harnesses horizontal projection technique for accurate skewness of writings precisely for above languages [38].

*Xichang Wang et al.* proposed an approach to clarify the difficulty of skew mechanism in the handwritten document image through the scanning action, a recent algorithm for skew angle detection stand on extreme gradient difference moreover Hough transform was suggested in the paper. First of all, maximum gradient difference was used to label text lines and then the method of non-text filtration to get rid of dots and smaller text lines. Experimental outcomes show that algorithm detects more rapid and accurate skew angles of document as compared to Hough transform alone [39].

*Yi-Feng Pan et al.* Purpose a hybrid approach for localize texts and robustly detect in scene images. Text detection and localization is important for content-based image analysis. This problem has some challenge due to the complex background, non-uniform illumination, and variations of text font, line orientation and variations of text size. A text region detector does help segment candidate text components by local binarization. ultimately, text constituent are grouped into text lines or words with a learning-based energy minimization method [40].

*Abhishek Arvind Gulhane et al.* proposed a fast technique to remove noise of an image which is mostly introduced due to environmental changes. Here focus on noise issues that is change image pixels value either off or on. The pixels are easily identified as noisy pixels in grayscale image but

difficult recognize in RGB 16 color image. Due to the reason of color combinations with white (pixel on) or black (pixel off) generate other color. They propose the technique that reduces noise in both RGB and grayscale image with recovery of original image [41].

*Youssef Bassil et al.* proposed a new post-processing method and algorithm for OCR error correction based on the “did you mean” spelling suggestion feature of Google’s online web search engine. The idea centers on using Google’s massive indexed data to detect and correct misspelled words in the OCR output text. The algorithm starts first by chopping the OCR text into several tokens of words. Then, each token is sent as a search query to Google’s search engine so that it gets processed. In case the query contains a misspelled word, Google will suggest a possible correction via its “did you mean” feature. Consequently, this spelling suggestion is to be considered as a correction for the misspelled query [42].

*Dhaval Salvi et al.* presented an approach that results the text subdivision with the extreme average plausibility for the resulting characters. A graph approach is used to describe the feasible locations for segmentation of proximate characters, and then an ordinary extensive path algorithm is implemented for identification of globally optimal segmentation. Smearing approaches use two RLSAs that are fuzzy and adaptive. The fuzzy RLSA measure is purposive for each pixel at the preliminary image and narrates that how up to a point one can see in case standing at a pixel along horizontal track. A gray scale image is now generated using this RLSA measure which is then binarized and text lines are extricated from the new generated image. The adaptive RLSA is a substitution to the traditional RLSA in the meaning that auxiliary smoothing constraints are set expressing geometrical characteristics of vicinal bridged components. The satisfaction of these constraints is continued by substitution of background pixels with foreground pixels [43].

*Saoi et al.* propose following stages: decomposing color image into RGB channel images and making 2D Wavelet Transform of each decomposed image, then using the unsupervised pixel block categorization with the k-means algorithm in incorporated feature vector space and integrating outputs of three channels by logical OR [44].

*Gllavata et al.* recommend a very precise method for text detection in video frames. It is used a slide window to move over the transformed wavelet images and characterized the zone with the allocation of high-frequency wave let coefficients. The author employ three region text, simple and complex backgrounds with k means algorithm [45].

*Angadi and Kodabagi* suggested a novel texture based text detection approach using high pass filtering in the DCT domain to arrest most of the background followed by, the feature vectors based on homogeneity and contrast are computed to identify text regions. The author mainly concentrates on the localization of rough estimate text blocks [46].

*Liu and Wang* proposed a novel approach for image binarization on grounds of robust threshold. Image binarization is evaluate as an optimization complication where the prime threshold could be launch out by partitioning binarization of Ostu threshold, mean while the approach of eliminating boundary effects is proposed to perfect the purpose of binarization optimization. The new approach is collated with a classic binarization method. The experimental result represented that the recent approach is preferable both, in keeping the authentic edge feature and available to more applications, particularly for the sake of superior effects of binarization optimization on images with affluent edge information [47].

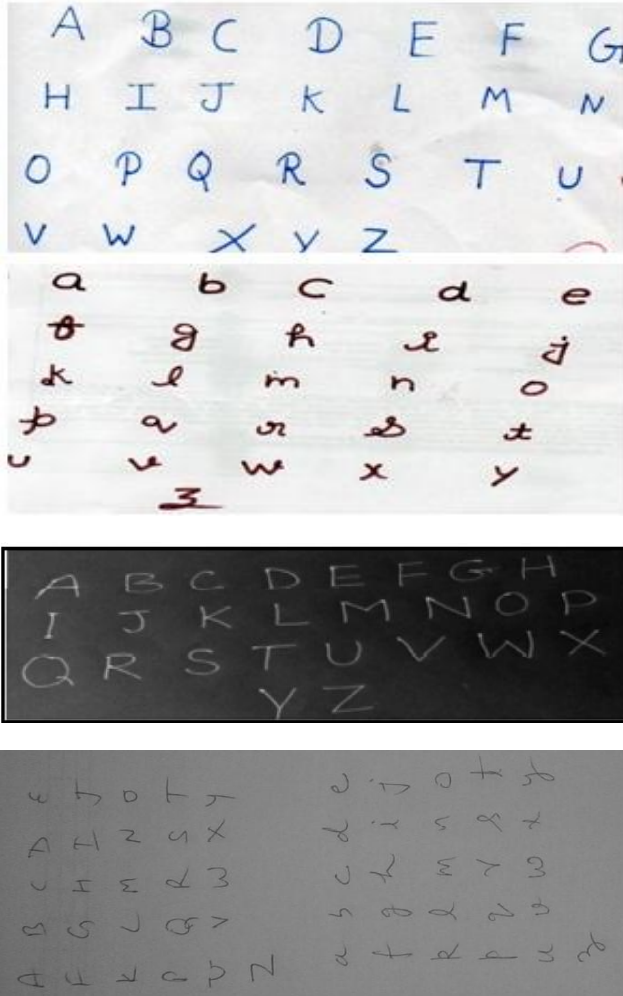


Shivakumara *et al.* proposed to extract CCs by executing K-means clustering in the Fourier- Laplacian domain, and finally use text straightness and edge density to detach false positives [48].

### 3. PHASES OF RECOGNITION

#### A. English Characters

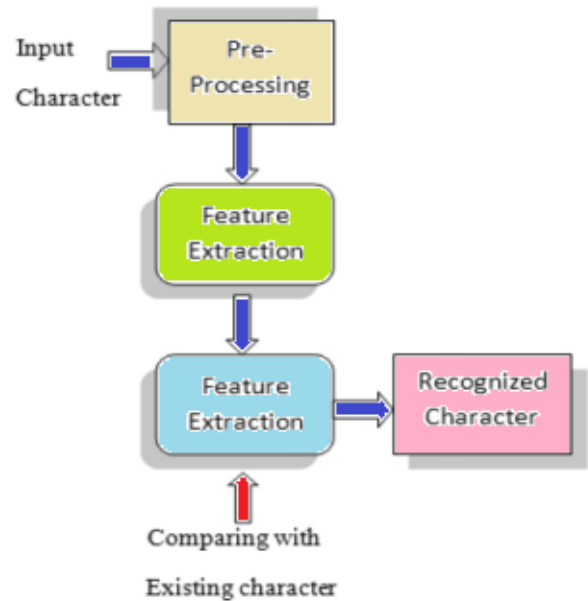
There are 26 characters in English language out of which 5 characters are vowels and 21 characters are consonants. A set of hand written English character are shown in Figure. 3



**Fig.3: A Set of Hand Written English Characters**

#### B. Different Phases of Handwritten Character Recognition

Each phase has been explained below (Figure 4):



**Fig. 4: Block Diagram of Character Recognition**

##### a. Pre-processing

Pre-processing is the name given to a family of procedures for smoothing, enhancing, Filtering, cleaning-up and otherwise massaging a digital image so that subsequent algorithm along the road to final classification can be made simple and more accurate. Various Pre-processing Methods are explained below:

##### (i) Binarization

Document image binarization refers to the conversion of a grayscale image into a binary image.

##### (ii) Noise Removal

The major objective of noise removal is to remove any unwanted bit-patterns, which do not have any significance in the output.

##### (iii) Skeletonization

Skeletonization refers to the thinning. Skeletonization is the process which reducing the width of an object from many pixels wide to just single pixel. This process can remove irregularities in letters and in turn, makes the recognition algorithm simpler because they only have to operate on a character stroke, which is only one pixel wide. It also reduces the memory space required for storing the information about the input characters and no doubt, this process reduces the processing time too.

##### (iv) Smoothing

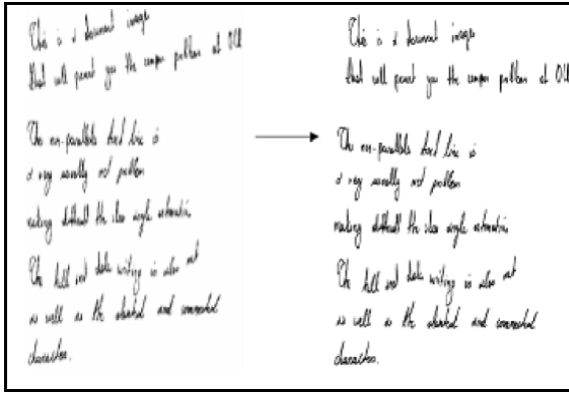
The objective of smoothing is to smooth contours of broken and/ or noisy input characters.

##### (v). Contour Smoothing

The objective of contour smoothing is to smooth contours of broken and/or noisy in Skewness input characters.

##### (vi) Skewness

Skewness introduces the Rotate in the bitmapped image of the scanned paper for character recognition system. It is normally occurs when the paper is not fed straight into the scanner. Most of the character recognition algorithms are sensitive to the orientation of the input document image, making it necessary to develop algorithms which can detect and correct the skew automatically (Figure 5).



**Fig. 5: Correction of Skewness**

After pre-processing phase, a cleaned image is obtained that goes to the segmentation phase.

#### b. Segmentation

It is a process that seeks to decompose an image into sub images of individual symbols. Character segmentation is the basic requirement that determines the utility of conventional Character Recognition systems. It includes line, word and character segmentation.

Different methods used can be classified based on the type of text and strategy being followed like recognition-based segmentation and cut classification method.

#### c. Feature extraction and Classification

Character Recognition system consists of two stages, feature extraction and classification. Feature extraction is the name given to a family of procedures for measuring the applicable shape information contained in a pattern so that the task of classifying the pattern is made easy by a formal procedure.

Classification stage is the basic decision making phase of the system and uses the features extracted in the earlier stage to identify the text segment according to preset rules.

Classification is apprehensive with making accord concerning the class membership of a pattern in question. A pattern class can be represented by a region or sub-space of the feature space. Classification then becomes a problem of determining the region of feature space in which an unknown pattern falls.

#### d. Post-processing

System results usually contain errors because of character classification and segmentation problems. For the correction of recognition errors, OCR systems apply contextual post-processing techniques. The two most common post-processing techniques for error correction are dictionary lookup and statistical approach. The advantage of statistical approach over dictionary-based methods is computational time and memory utilization. The simplest way of incorporating the context information is the utilization of a dictionary for correcting the minor mistakes (Figure 6).

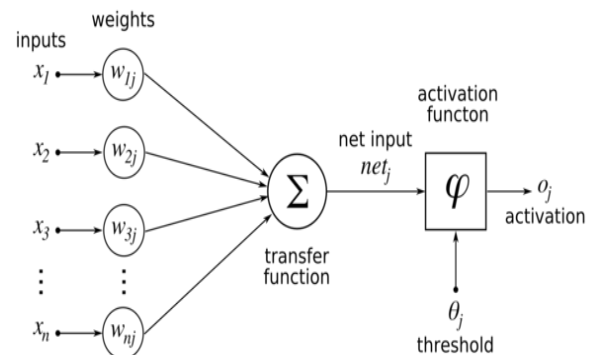


**Fig. 6: Post Processing**

## 4. ARTIFICIAL NEURAL NETWORK

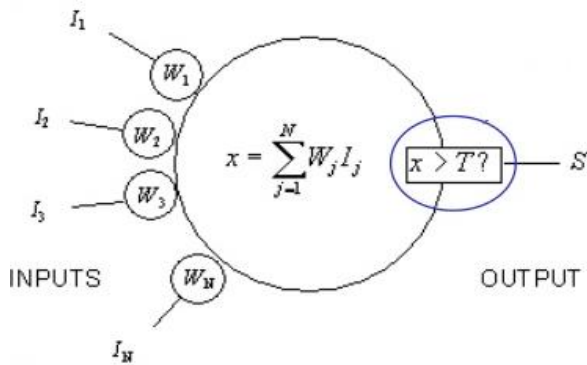
Basically artificial neural networks are biologically inspired networks. In human brain there are millions of tiny processors know as neurons connected to each other by means of synapses. These neurons forward the information in form of chemical signals to other neurons on the basis of those signals which they receive. The neuron which tries to mimic the behavior of biological neuron is known as artificial neuron. Initially neuron model was proposed by McCulloch and Pitts in the year 1943 after 20 years of research. The main motive behind the research of neural network is a keen desire to create a machine that works similar to the manner human brain works. In different areas, the solution of a broad spectrum of problems can be gained by neural networks [49-51].

Computers treat the characters like letters or digits as binary graphics while human brains can identify and memorize them. Hence, algorithms are necessary key factors for the identification and recognition of each character [52]. The neuron model which is used in ANNs is illustrated in Figure 7 [53]. For a neuron, the inputs  $x_i$  are weighted by weights  $w_i$  and both are summed together. If this sum is more than a threshold, the neuron is active.



**Fig. 7. Single neuron "Perceptron" model**

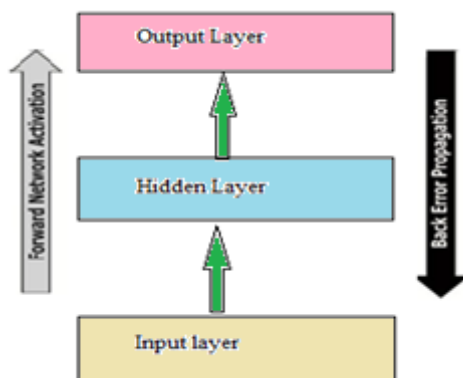
The below Figure 8, is a mathematical representation of the Artificial Neural Network (ANN). Neural network recognizers regulate from an early image training set. The trained network then reflects the character recognition. Each neural network individually learns the characteristics that distinguish training images followed by looking for equivalent characteristics in the desired image to be identified. Although, neural networks can be erroneous if they learn attributes that are not salient in the focused data, NNs can be setup promptly. It may be concluded if neural networks are not accordingly trained then results are unreliable.



**Fig 8. Mathematical Representation of ANN**

#### A. Backpropagation Algorithm

An Error Backpropagation (EBP) or simply Backpropagation (BP) neural network consists of at least three layers of units, input layer, intermediate hidden layer and an output layer. When algorithm is cycled, an input pattern is propagated forward to the output units through the intervening input-to-hidden and hidden-to-output weights [54]. The working of a BP algorithm is represented by Figure 9. The iteratively processing a data set of prepare values, comparing the network's declare for each set with the actual target values [55]. For each instruct dataset, the weights are modified so as to minimize the mean squared error between the network's prediction and the target value [56, 57]. These modifications are made in the "backwards" direction. Although it is not decided, in ordinary the weights will ultimately converge, and the learning process barricade. Neural network uses BP algorithm which is an approach and a supervised algorithm that learns by first computing the output applying a feed forward network, then calculating the error signal and fecundate the error backwards through the network [58]. The BP algorithm's most unambiguous feature is the error that the neural network gets on its output. In BP algorithm, the authors look at the weight/change, which mostly decrease the error from the output [59]. This algorithm is a procedure that depends on the ascent value of the moment. The learning starts when all of the training data was showed to the network at least once. The learning method, as for each network learning algorithm, consists of the alteration of the weights.

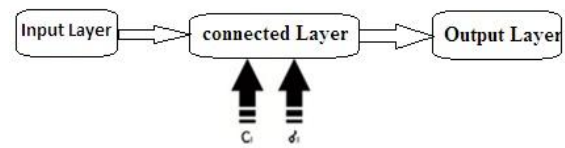


**Fig. 9. Working of Back propagation algorithm**

#### B. RBFNN

The RBFNN is composed of three layers: an input layer, a concealed (connected) layer and an output layer as represented by Figure 10. The input layer corresponds to the input vector space and the output layer corresponds to the

classifications [60]. The concealed layer is considered as radial basis function nodes. The basic function used for this NN is Gauss kernel function.



**Fig. 10. RBFNN Structure**

#### C. Parallel BP Network

Among many neural networks arithmetic in character recognition, BP arithmetic is the most robust one. In this paper, a parallel BP neural network is used to recognize characters. a simple BP neural network is not efficient enough in character recognition with a recognition rate of less than 90%. The direct reason is the loss of spatial compounding information. The parallel BP neural network consists of two simple BP neural networks: BP neural networks A and BP neural networks B. In the input layer of neural networks A, the character image is input in a row-first way so that the row spatial information of the character is preserved. And in the input layer of neural networks B, the character image is input in a column-first way so that the column spatial information is preserved. Therefore, the proposed parallel BP neural networks can preserve the spatial compounding information and can effectively improve the recognition rate.

#### D. Feed Forward Artificial Neural Network (FF-ANN)

Feed-forward neural networks (FFANN) are highlighted among well accepted structures among the series of artificial neural networks. These coherent networks are broadly used to solve complicated issues by modeling gordian input-output relationships. Nonetheless, FFANNs frequently end up being over trained. They adopt trials-and-errors for the sake of feasible mileage of parameters for convergence of the global optimal. The learning procedure of an FFANN cannot guarantee the global ideal, occasionally trapping the network into the local optimal. The local optimal. The back-propagation learning algorithm (BPLA) is a broadly used approach for FFANN learning in many fields. It has the sizeable convenience of elementary implementation.

## 5. CONCLUSION

There is a noteworthy development in the character recognition methods during last few years. In the related fields like face recognition and image recognition, a range of techniques have emerged and influenced by developments. This review has proposed an amalgamation of these methods under some basic strategies. Hence there is an opportunity that this comprehensive discussion will provide observation into the concepts involved, and perhaps provoke more advances in the area. Nature and quality of the text are the two main difficulties to perform accurate recognition. Generally, abnormal segmentation rates for immoral material increase progressively from machine print to handprint to soft writing.

Current research exploit models have characters, words and phrases and entire documents and powerful tools such as HMM. While this review have focused on the recognition problem and it is clear that segmentation and classification have to be treated in an integrated manner to obtain huge reliability in complicated cases. It is considered that brilliant use of features and NNs have boost the improved accuracies. A character can be classified on the basis of features of each character that are necessarily required. In case if the exact



pattern is not available in the database, neural network support the system to locate and recognize the characters. The matching consider high speed, but it is not very effective in case of font defilement, font slant, font discrepancy, stroke breaking and stroke connection because of the instrument itself and milieu. It may be accomplished that the use of ANN based methods and categorization will be more potent with enhanced processing speed and exactness. The advantage of ANN is that they are more similar to a real living nervous system. In case of simultaneous satisfaction of multiple constraints, the parallel organizations of the system can give flexible solutions to various problems. Neural Network is also easier to handle and insensible to noise, because of less involvement of human work as compared to classical statistical analysis. ANN has the talent to recover smoothly from distortion in the input data and their efficiency of learning. They are often excellent at solving problems that are much complex for conventional technologies. For example, neural networks are capable to solve problems that an algorithmic solution is too complex to be found and/or do not have an algorithmic solution. They have few unfavorable points because of their response to anything and disappearance of change in output when small changes in the input normally have been made. The parallel architecture of ANNs is also beneficial to achieve high computational rates. The ANN's have broad scope to use in various areas like Air Fraud Detection, Machine Control, Traffic Control, Music Composition, Medical Diagnosis, Images and Finger Prints Recognition, Staff Scheduling, Handwriting Recognition, Voice Recognition and Weather Prediction etc. The synthesis of more than two techniques to improve the accuracy of the system may also be possibly an icing on the cake. Lists of references also have been included in order to provide more-detailed information of the approaches described.

## 6. ACKNOWLEDGMENTS

Authors are grateful to Prof. R M Dubey, Vice Chancellor of IFTM University, Moradabad for providing necessary facilities for this research work.

## 7. DECLARATION OF INTEREST

The authors declare that they have no competing interests.

## 8. REFERENCES

- [1] Sushil Gangwar, Krishan Kumar, "3D Face Recognition Based On Extracting PCA Methods", *International Journal of Engineering Research and Applications (IJERA)*, ISSN: 2248-9622, Vol2, Issue 2, Mar-Apr 2012, pp 693-696.
- [2] Amit Choudhary, "A Review of Various Soft Computing Techniques in the Domain of Handwriting Recognition", *International Journal of Information & Computation Technology*. Vol. 4, No. 6, pp. 601-606.
- [3] Rahul Malhotra, Narinder Singh and Yaduvir Singh. Soft computing techniques for process control applications. *International Journal on Soft Computing (IJSC)*, Vol.2, No.3 (2011).
- [4] A. Thilagavathy, K. Aarthi and A. Chilambuchelvan. Text detection and extraction from videos using ann based network. *International Journal on Soft Computing, Artificial Intelligence and Applications*. Vol. 1, No.1 (2012), pp.19-28.
- [5] Yi-Feng Pan, Xinwen Hou, and Cheng-Lin Liu. A robust system to detect and localize texts in natural scene images. In *International Workshop on Document Analysis Systems*, 2008.
- [6] Yi-Feng Pan, Xinwen Hou, and Cheng-Lin Liu. Text localization in natural scene images based on conditional random field. In *ICDAR*, 2009.
- [7] P. M. Kamble and R. S. Hegadi, "Handwritten marathi basic character recognition using statistical method," in *Emerging Research in Computing, Information, Communication and Applications*. Elsevier, 2014, Vol. 3, 2014, pp. 28–33.
- [8] Adam Coates, Blake Carpenter, Carl Case, Sanjeev Satheesh, Bipin Suresh, Tao Wang, David J. Wu, and Andrew Y. Ng. Text detection and character recognition in scene images with unsupervised feature learning. In *ICDAR*, 2011.
- [9] V. Sagar and K. Kumar, "A symmetric key cryptography using genetic algorithm and error back propagation neural network," *Computing for Sustainable Global Development (INDIACom)*, 2015 2nd International Conference on, New Delhi, 2015, pp. 1386-1391.
- [10] Vikas sagar, Krishan Kumar "A Symmetric Key Cryptography Using Counter Propagation Neural Network", *International Conference on Information and Communication Technology for Competitive Strategies*, *ACM-ICPS Proceedings Volume ISBN No 978-1-4503-3216-3*.
- [11] D. C. Ciresan, U. Meier, and J. Schmidhuber. Multi-column deep neural networks for image classification. Technical Report IDSIA-04-12, Dalle Molle Institute for Artificial Intelligence, 2012.
- [12] Adam Coates, Blake Carpenter, Carl Case, Sanjeev Satheesh, Bipin Suresh, Tao Wang, David J. Wu, and Andrew Y. Ng. Text detection and character recognition in scene images with unsupervised feature learning. In *ICDAR*, 2011.
- [13] Chucai Yi and Yingli Tian. Text Detection in Natural Scene Images by Stroke Gabor Words. 2011 International Conference on Document Analysis and Recognition. 177-181.
- [14] Xu-Cheng Yin, Xuwang Yin, Kaizhu Huang, and Hong-Wei Hao. Robust Text Detection in Natural Scene Images. 2013.
- [15] Parshuram M. Kamble, Ravinda S. Hegadi. Handwritten Marathi character recognition using R-HOG Feature. *Procedia Computer Science* 45 (2015) 266 – 274.
- [16] Yi-Feng Pan, Xinwen Hou, Cheng-Lin Liu. Text Localization in Natural Scene Images based on Conditional Random Field. 2009 10th International Conference on Document Analysis and Recognition.
- [17] A. Thilagavathy, K. Aarthi, A. Chilambuchelvan. Text Detection and Extraction From Videos Using ANN Based Network. *International Journal on Soft Computing, Artificial Intelligence and Applications (IJSCAI)*, Vol.1, No.1, August 2012.
- [18] Xiangrong Chen and A.L. Yuille. Detecting and reading text in natural scenes. In *Computer Vision and Pattern Recognition*, volume 2, 2004.
- [19] Marc'Aurelio Ranzato, Christopher Poultney, S. Chopra, and Y. LeCun. Efficient learning of sparse

- representations with an energy-based model. In NIPS, 2007.
- [20] B. Epshtein, E. Ofek, and Y. Wexler. Detecting text in natural scenes with stroke width transform. In CVPR, 2010.
- [21] T. M. Rath and R. Manmatha: Features for Word Spotting in Historical Manuscripts. In: Proc. of the 7th Int'l Conf. on Document Analysis and Recognition (ICDAR), Edinburgh, Scotland, August 3-6, 2003, vol. 1, pp. 218-222.
- [22] M.K. Jindal, R.K. Sharma, G.S. Lehal, "Segmentation of Horizontally Overlapping lines in Printed Gurmukhi Script", IEEE, 2006.
- [23] A. Zahour, B. Taconet, L. Sulem, and W. Boussellaa, "Overlapping and multi-touching text line segmentation by Block Covering analysis," Pattern Analysis and Applications, Vol. 12, pp. 335-351, 2008.
- [24] G. Louloudis, et al. Text line and word segmentation of hand written documents. Pattern Recognition, 42, 3169—3183, 2009.
- [25] N Stamatopoulos, B Gatos, I Pratikakis, SJ Perantonis. Goal-oriented rectification of camera-based document images. Image Processing, IEEE Transactions on 20 (4), 910-920, 2011.
- [26] Yi Li, Yefeng Zheng, David Doermann, Stefan Jaeger," Script-Independent Text Line Segmentation in Freestyle Handwritten Documents." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 8, Aug.2008.
- [27] T Lambrianidis, K Lyroudia, O Pandelidou, A Nicolaou. Evaluation of periapical radiographs in the recognition of C-shaped mandibular second molars. Int Endod J 2001 Vol 34 (458-62)
- [28] A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, "Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks," in ICML, Pittsburgh, USA, 2006.
- [29] F. Ahmed and S. Farid, "Application of Niblack's Method on Images," International Conference on Emerging Technologies, 2009.
- [30] F. Yin, C.L. Liu, Handwritten text line extraction based on minimal spanning tree clustering, Proc. 5th Int. Conf. on Wavelet Analysis and Pattern Recognition, Vol.3, pp. 1123- 1128, 2007.
- [31] Xiaojun Du, Wumo Pan, Tien D. Bui," Text line segmentation in handwritten documents using Mumford–Shahmodel," Pattern Recognition vol. 42, pp. 3136 – 3145, 2009.
- [32] A. Suliman, A. Shakil, M. N. Sulaiman, M. Othman, R. Wirza, Hybrid of HMM and Fuzzy Logic for handwritten character recognition DOI: 10.1109/ITSIM.2008.4631674 Conference: Information Technology, 2008. ITSIM 2008. International Symposium on, Volume: 2
- [33] Vassilis Papavassilioua, Themis Stafylakis, Vassilis Katsouros, George Carayannis, Handwritten document image segmentation into text lines and words, Pattern Recognition 43 (2010) 369—377.
- [34] Alireza Alaei, P. Nagabhushan, Umapada Pal. A New Dataset of Persian Handwritten Documents and its Segmentation. 2011 7th Iranian Conference on Machine Vision and Image Processing, MVIP2011-Proceedings 01/2011; DOI: 10.1109/IranianMVIP.2011.6121553.
- [35] Brijmohan Singh, Ankush Mittal, M.A. Ansari, Debashis Ghosh. Handwritten Devanagari Word Recognition: A Curvelet Transform Based Approach. International Journal On Computer Science And Engineering (IJCSE) Vol. 3 No. 4 ISSN : 0975-3397 40634.
- [36] K. B. M. R. Batuwita, G. E. M. D. C. Bandara. An Improved Segmentation Algorithm for Individual Offline Handwritten Character Segmentation. CIMCA '05 Proceedings of the International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce Vol-2 (CIMCA-IAWTIC'06) - Volume 02, 982-988, 2005.
- [37] <https://books.google.co.in/books?isbn=3662485702>.
- [38] Mohammad Abu Obaida, Tanay Kumar Roy, Md. Abu Horair, Md. Jakir Hossain. Skew Correction Function of OCR: Stroke-Whitespace based Algorithmic Approach. International Journal of Computer Applications (0975 – 8887), Volume 28– No.8, August 2011.
- [39] Bikash Shaw, Swapna Kr. Parui, Malayappan Shridhar. Offline Handwritten Devanagari word Recognition: A Holistic Approach Based On Directional Chain Code Feature And HMM. International Conference On Information Technology IEEE ICIT 978-0-7695-3513-5/08 2008.
- [40] Yi-Feng Pan, Xinwen Hou, Cheng-Lin Liu. A Hybrid Approach to Detect and Localize Texts in Natural Scene Images. Chinese Academy of Sciences (CASIA), 20(3):800-13. DOI: 10.1109/TIP.2010.2070803.
- [41] Abhishek arvind gulhane. Noise Reduction of an Image by using Function Approximation Techniques. 2709 (2004).
- [42] Youssef Bassil, Mohammad Alwani, Ocr post-processing error correction algorithm using google's online spelling suggestion. Journal of Emerging Trends in Computing and Information Sciences, ISSN 2079-8407, Vol. 3, No. 1, January 2012.
- [43] Dhaval Salvi, Jun Zhou, Jarrell Waggoner, and Song Wang. Handwritten Text Segmentation using Average Longest Path Algorithm. Ijarcse, Vol. 3(5), 2008.
- [44] T.Saoi, H. Goto, H. Kobayashi, Text detection in color scene images based on unsupervised clustering of multi-channel wavelet features, in: Eighth International Conference on Document Analysis and Recognition (ICDAR'05), vol.2, 2005, pp. 690–694.
- [45] S.A. Angadi, M.M. Kodabagi, A texture based methodology for text region extraction from low resolution natural scene images, in: Advance Computing Conference, 2010, pp. 121–128.
- [46] J. Gllavata, R. Ewerth, B. Freisleben, Text detection in images based on unsupervised classification of high-frequency wavelet coefficients, in: Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, IEEE, vol.1, 2004, pp. 425–428.

- [47] J. Liu, C. Wang, An algorithm for image binarization based on adaptive threshold, in: 2009 Chinese Control and Decision Conference, CCDC'09, IEEE, 2009, pp. 3958–3962.
- [48] P. Shivakumara, T. Phan, C. Tan, A Laplacian approach to multi-oriented text detection in video, *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (2) (2011) 412–419.
- [49] Jerod J. Weinman, Erik Learned-Miller, and Allen R. Hanson. A discriminative semimarkov model for robust scene text recognition. In *Proc. IAPR International Conference on Pattern Recognition*, Dec. 2008.
- [50] L. Neumann and J. Matas. A method for text localization and recognition in real-world images. In *AACCV*, 2010.
- [51] K. Wang, B. Babenko, and S. Belongie. End-to-end scene text recognition. In *ICCV*, 2011.
- [52] Attaullah Khawaja, Shen Tingzhi, Noor Mohammad Memon, AltafRajpa, “Recognition of printed Chinese characters by using Neural Network”, 1-4244-0794-X/06/\$20.00 ©2006 IEEE, pp 169-172.
- [53] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. Wiley, 2001.
- [54] Jerod Weinman, Erik Learned-Miller, and Allen R. Hanson. Scene text recognition using similarity and a lexicon with sparse belief propagation. In *Transactions on Pattern Analysis and Machine Intelligence*, volume 31, 2009.
- [55] E.Kavallieratos, N.Antoniades, N.Fakotakis and G.Kokkinakis, “Extraction and recognition of handwritten alphanumeric characters from application forms”.
- [56] X. Fan and G. Fan. Graphical Models for Joint Segmentation and Recognition of License Plate Characters. *IEEE Signal Processing Letters*, 16(1), 2009.
- [57] Jerod Weinman, Erik Learned-Miller, and Allen R. Hanson. Scene text recognition using similarity and a lexicon with sparse belief propagation. In *Transactions on Pattern Analysis and Machine Intelligence*, volume 31, 2009.
- [58] Yuk Yirtg Chung, M.,an To Wong, “Handwritten Character Recognition By Fourier Descriptors And Neural Network”, 1997 IEEE TENCON, pp 391-394.
- [59] Rókus Arnold, Póth Miklós, “Character Recognition Using Neural Networks”, *CINTI 2010*, 978-1-4244-9280-0/10/\$26.00 ©2010 IEEE, 311-314.
- [60] Suruchi G. Dedgaonkar, Anjali A. Chandavale, Ashok M. Sapkal. Survey of Methods for Character Recognition. *International Journal of Engineering and Innovative Technology (IJEIT)*, Volume 1, Issue 5, May 2012.