



A LITERATURE REVIEW ON HAND WRITTEN CHARACTER RECOGNITION

MANSI SHAH AND GORDHAN B JETHAVA

Department of Computer Science & Engineering Parul Institute of Technology, Gujarat, India.

Information Technology Department Parul Institute of Engg. & Technology, Gujarat, India.

ABSTRACT:

Nowadays character recognition has gained lot of attention in the field of pattern recognition due to its application in various fields. Optical Character Recognition (OCR) and Handwritten Character Recognition (HCR) has specific domain to apply. OCR system is most suitable for the applications like multi choice examinations, printed postal address resolution etc. While application of HCR is wider compare to OCR. HCR is useful in cheque processing in banks; almost all kind of form processing systems, handwritten postal address resolution and many more. In coming days, character recognition system might serve as a key factor to create paperless environment by digitizing and processing existing paper documents. In this paper, we have provided the detail study on existing methods for handwritten character recognition.

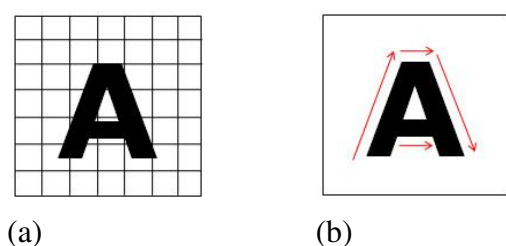
KEYWORDS:

OCR, HCR, Features, Training, classification.

I. INTRODUCTION

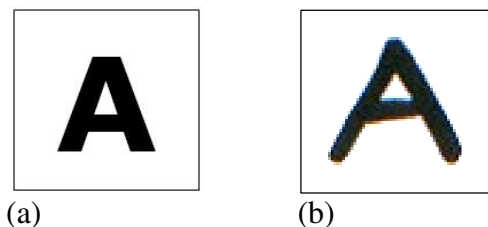
Character recognition is an art of detecting segmenting and identifying characters from image. More precisely *Character recognition is process of detecting and recognizing characters from input image and converts it into ASCII or other equivalent machine editable form* [1], [2], [3]. It contributes immensely to the advancement of automation process and improving the interface between man and machine in many applications [4]. Character recognition is one of the most interesting and fascinating

areas of pattern recognition and artificial intelligence [5], [6]. Character recognition is getting more and more attention since last decade due to its wide range of application. Conversion of handwritten characters is important for making several important documents related to our history, such as manuscripts, into machine editable form so that it can be easily accessed and preserved [7]. Lots of independent work is going on in Optical Character Recognition that is processing of printed/computer generated document and handwritten and manually created document processing i.e. handwritten character recognition. Character recognition process can be classified in two categories. Offline character recognition (Fig. 1(a)) and online character recognition (Fig. 1(b)).



**Fig 1. (a). Offline Character Recognition,
(b). Online Character Recognition**

In offline character recognition system, document is first generated, digitized, stored in computer and then it is processed. While in case of online character recognition system, character is processed while it was under creation. External factors like pressure speed of writing, stroke making etc does not have any influence in case of offline system but they have great impact on online system. Again, offline or online system can be applied on optical character (Fig 2.(a)) or handwritten characters (Fig 2.(b)). Based on that system can be classified as OCR or HCR respectively. The online methods have been superior to their counterpart i.e. offline methods due to temporal information present in character generation [4].



**Fig 2. (a). Optical Character, (b).
Handwritten Character**

Accuracy of HCR is still limited to 90 percent due to presence of large variation in shape, scale, style, orientation etc. [8]. Character processing systems are domain and application specific, like there it

is not possible to design generic system which can process all kind of scripts and language. Lots of work has been done on European languages and Arabic (Urdu) language. Whereas domestic languages like Hindi, Punjabi, Bangla, Tamil, Gujarati etc. are very less explored due to limited usage. In this paper, our focus is to carry out in depth literature survey on handwritten character recognition methods.

Rest of the paper is organized as follow. Section II describes the state of art followed by detail survey on existing handwritten character recognition techniques. Section IV describes the conclusions followed by references in next section.

II. STATE OF ART

A. Working Principle

Any character recognition system goes under following steps, i.e. Image acquisition, Preprocessing, Segmentation, Feature extraction, classification and post processing. Block diagram is of general character recognition system is shown in fig 3.

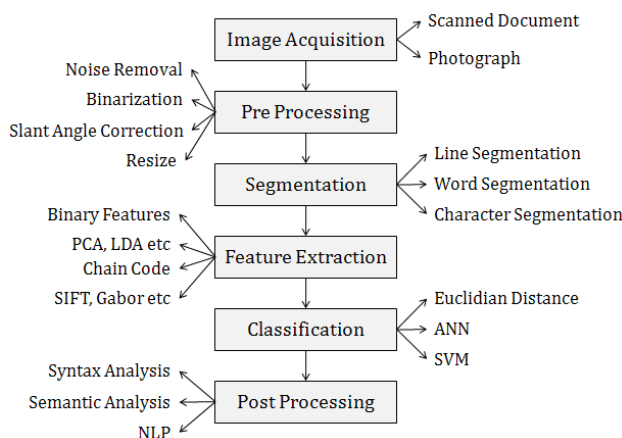


Fig 3. Block Diagram of Character Recognition System

Image acquisition: Images for HCR system might be acquired by scanning handwritten document or by capturing photograph of document or by directly writing in computer using stylus. This is also known as digitization process.

Preprocessing: Preprocessing involves series of operations performed to enhance to make it suitable for segmentation [4]. Preprocessing step involves noise removal generated during document generation. Proper filter like mean filter, min-max filter, Gaussian filter etc may be applied to remove noise from document. Binarization process converts gray scale or colored image to black and white image. Binary morphological operations like opening, closing, thinning, hole filling etc may be applied to enhance

visibility and structural information of character. If document is scanned then it may not be perfectly horizontally aligned, so we need to align it by performing slant angle correction. Input document may be resized if it is too large in size to reduce dimensions to improve speed of processing. However reducing dimension below certain level may remove some useful features too.

Segmentation: Generally document is processed in hierarchical way. At first level lines are segmented using row histogram. From each row, words are extracted using column histogram and finally characters are extracted from words. Accuracy of final result is highly depends on accuracy of segmentation.

Feature Extraction: Feature extraction is the heart of any pattern recognition application. Feature extraction techniques like Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA), Chain Code (CC), Scale Invariant Feature Extraction (SIFT), zoning, Gradient based features, Histogram might be applied to extract the features of individual characters. These features are used to train the system.

Classification: When input image is presented to HCR system, its features are extracted and given as an input to the trained classifier like artificial neural network or support vector machine. Classifiers compare the input feature with stored pattern and find out the best matching class for input.

Post processing: This step is not compulsory; sometimes it helps to improve the accuracy of recognition. Syntax analysis, semantic analysis kind of higher level concepts might be applied to check the context of recognized character.

B. Classifiers

Taxicab Metric: Simplest measure is the absolute difference of two patterns. Taxicab distance between two points is the sum of absolute differences of their coordinates. Let's say A is the template feature image which is to be compared with test image Y. Assume that image is of M X N. Taxicab measure between two images A and B is given as follow:

$$M_d = \sum_{i=1}^M \sum_{j=1}^N |A(i, j) - B(i, j)|$$

Taxicab metric is also known as rectilinear distance, L_1 distance, city block distance, Manhattan distance etc.

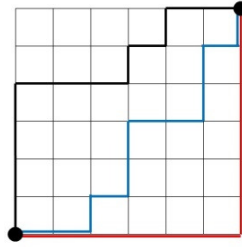


Figure 4. Different ways to measure Manhattan distance

Euclidian distance / L2 Norm: Euclidean distance is widely used distance measure, which is also known as L2 norm or nearest neighbor classifier. Euclidean distance is ordinary distance between two points which could be measured using ruler (Fig. 5).

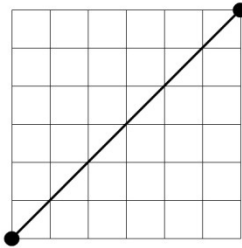


Figure 5. Euclidean distance between two points

Euclidean distance between two vectors of size N is given by:

$$ED = \sqrt{\sum_{i=1}^N (A_i - B_i)^2}$$

In same way, Euclidean distance between two dimensional images can be given by:

$$ED = \sqrt{\sum_{i=1}^M \sum_{j=1}^N (A(i, j) - B(i, j))^2}$$

This equation can be generalized for any dimensions and any norm. L norm of one dimensional vector can be given as:

$$ED_L = \sqrt[L]{\sum_{i=1}^N (A_i - B_i)^L}$$

When L is 2, it becomes Euclidean measure and hence Euclidean distance is also known as L2-norm. Different types of Nearest Neighbor techniques have been reported in the literature [9].

Chess board distance: Chess board distance between two vectors is the greatest of their differences along any coordinate dimension. It is also known as L_{∞} metric or chebyshev distance. Chess board distance between two vectors X and Y with n coordinates,

$$d_{CB} = \max_i (|X_i - Y_i|)$$

3	3	3	3	3	3	3
3	2	2	2	2	2	3
3	2	1	1	1	2	3
3	2	1	1	1	2	3
3	2	1	1	1	2	3
3	2	2	2	2	2	3
3	3	3	3	3	3	3

Figure 6. Chess board distance

Cross correlation: Cross correlation is similarity measure between two signals in signal processing. It also has the applications in pattern recognition to compare two patterns or to find distance or similarity between two patterns. The correlation coefficient is one of the popular metric used in the literature to provide comparison of two images. And the correlation coefficient is given as,

$$R = \frac{\sum_{i=1}^M \sum_{j=1}^N (A_{ij} - \bar{A}) \cdot (B_{ij} - \bar{B})}{\sqrt{\left(\sum_{i=1}^M \sum_{j=1}^N (A_{ij} - \bar{A})^2\right) \left(\sum_{i=1}^M \sum_{j=1}^N (B_{ij} - \bar{B})^2\right)}}$$

Where R indicates the Correlation Coefficient between two matrices A and B, where A and B are matrices or vectors of the same size. \bar{A} and \bar{B} are mean of A and B respectively. A correlation coefficient is a numerical, descriptive measure of the strength of the linear relationship between two variables. Values for the correlation coefficient range between -1 and +1; with a correlation coefficient of +1 indicating that the two variables have a perfect, upward-sloping (+) linear relationship and a correlation coefficient of -1 showing that the two variables are perfectly related in a downward-sloping, (-) linear sense. A correlation coefficient of 0 demonstrates that the variables have no relationship, and are independent

Artificial Neural network (ANN): Artificial neural network is widely accepted classifier for diverse patterns. ANN works on phenomenon of biological neurons and learns to classify unseen data. Multilayer neural networks have been widely used in pattern recognition applications. Various paradigms have been used. The different network models are specified by:

Network topology: the number of neurons and how the neurons are interconnected.

Node characteristics: the type of non-linear transfer function used by the neuron for calculating the output value.

Training rules: specify how the weights are initially set and adjusted to improve performance of the network.

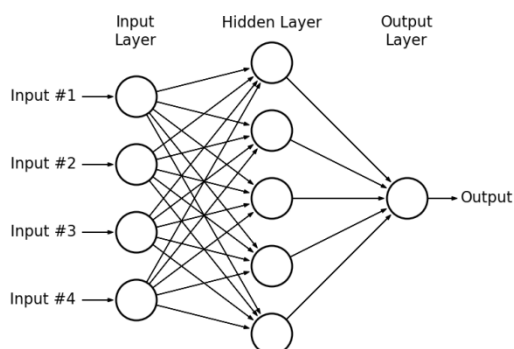


Figure 7: Architecture of Multilayer neural network

Architecture of neural network depends on nature and complexity of applications. However multilayer neural network with proper choice of parameter is capable enough to classify almost any pattern. The back propagation model or multi-layer perceptron is a neural network that utilizes a supervised learning technique. Typically there are one or more layers of hidden nodes between the input and output nodes.

There are so many parameters that control the performance of neural network, like

Number of layers

Number of neurons in each layer

Transfer function used between two layers

Learning algorithm

Number of epochs

*Support vector machine (SVM):*Support vector machine is supervised learning tool, which is used for classification and regression. The basic SVM takes a set of input data and predicts, for each given input. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. More formally, a support vector machine constructs a hyper plane or set of hyper planes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyper plane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.

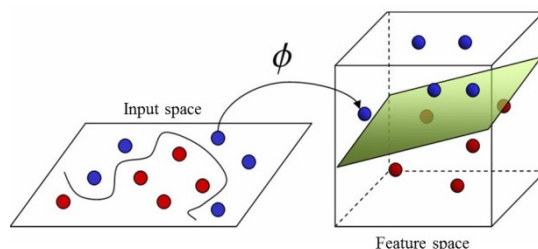


Figure 8. Feature transformation

Whereas the original problem may be stated in a finite dimensional space, it often happens that the sets to discriminate are not linearly separable in that space. For this reason, it was proposed that the

original finite-dimensional space be mapped into a much higher-dimensional space, presumably making the separation easier in that space.

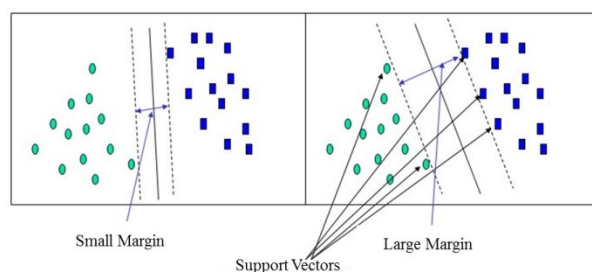


Figure 9. Maximum margin classification by SVM

III. LITERATURE REVIEW

Image processing and pattern recognition plays significant role in handwritten character recognition. In [10], Rajbala et al have discussed various types of classification of feature extraction methods like statistical feature based methods, structural feature based methods and global transformation techniques. Statistical methods are based on planning of how data are selected. It uses the information of statistical distribution of pixels in image. It can be mainly classified in three categories: 1). Partitioning in regions, 2). Profile generation and projections 3). Distances and crossing. Structural features are extracted from structure and geometry of character like number of horizontal and vertical lines, aspect ratio, number of cross points, number of loops, number of branch points, number of strokes, number of curves etc. Global transformation features are calculated by converting image in frequency domain like Discrete Fourier Transformation (DFT), Discrete Cosine Transformation (DCT), Discrete Wavelet Transformation (DWT), Gabor filtering, Walsh-Hadamard transformation etc.

Feature extracted can be either low level or high level. Low level features include width, height, curliness, aspect ratio etc of the character. These alone cannot be used to distinguish one character from another in the character set of the language [11]. So, there are a number of other high level features which include number and position of loops, straight lines, headlines, curves etc.

Tirthraj Dash et al have discussed HCR using associative memory net (AMN) in their paper [12]. They have directly worked at pixel level. Dataset was designed in MS Paint 6.1 with normal Arial font of size 28. Dimension of image was kept 31 X 39. Once characters are extracted, their binary pixel values are directly used to train AMN.

I.K.Pathan et al have proposed offline approach for handwritten isolated Urdu characters in their work mentioned in [13]. Urdu character may contain one, two, three or four segment. In which one component is known as primary (generally represents large continuous stroke) and rest of all are known as secondary components (generally represents small stroke or dots). Authors have used moment

invariants (MI) feature to recognize the characters. MI features are well known to be invariant under rotation, translation, scaling and reflection. MI features are measure of the pixel distribution around the center of gravity of character and it captures the global character shape information. If character image is single component than it is normalized in 60×60 pixels and horizontally divided in equal 3 parts. 7 MI are extracted from each zone and 7 MI are calculated from overall image, so total of 28 features are used to train SVM. And if image is having multi component than 28 MI are extracted from primary component (60×60) and 21 MI are extracted from secondary component (22×22). Separate SVM are trained for both and decision is taken based on rules satisfying some criteria. Proposed system claim to get highest 93.59 % accuracy.

In [4], Pradeep et al have proposed neural network based classification of handwritten character recognition system. Each individual character is resized to 30×20 pixels for processing. They are using binary features to train neural network. However such features are not robust. In post processing stage, recognized characters are converted to ASCII format. Input layer has 600 neurons equal to number of pixels. Output layer has 26 neurons as English has 26 alphabets. Proposed ANN uses back propagation algorithm with momentum and adaptive learning rate.

Rajib et al have proposed Hidden Markov Model based system for English HCR in their literature [8]. They have employed global as well as local feature extraction methods. Global feature involves four gradient features, six projection features and four curvature features. And to extract local features, image is divided in to nine equal blocks and 4 gradient features are calculated from each block, so total of 36 features are extracted. So overall feature vector contains 50 features per character. $O = [G(4) P(6) C(4) L(36)]$, where G, P, C and L represents global gradient, projection, curvature and local gradient features respectively. Number in parenthesis represents number of respective features. HMM is trained using these feature and experiment is carried out. Post processing is also applied after recognition phase of HMM to highly confused group of characters like N and M, O and Q, C and O etc. For each group new feature is calculated to discriminate characters within the group.

Gradient features based method is discussed in [14] by Ashutosh et al. Experiment is carried out on Hindi, third most popular language in the world. The first research work on handwritten Devnagari characters was published in 1977. 300 Million People use the Devnagari script for documentation in central and north region of India [8]. In proposed method, Gradient Vector is calculated at each pixel and image is divided in 9×9 blocks. Then strength of gradient is accumulated in eight standard directions in each sub block. 9×9 blocks is further down sampled to 5×5 block using Gaussian filter. Preprocessing steps for proposed methods are as follows: Intensity values of image were adjusted and then images were converted to binary with threshold value 0.8. Connected component with pixel density less than 30 were removed from further processing. Median filter was applied to remove pepper and salt noise present in binary images. And finally, individual characters were extracted by row and column histogram processing and normalized to 90×90 pixel block. Gradient feature were extracted using sobel operator.

1	2	1	-1	0	1
0	0	0	-2	0	2
-1	-2	-1	-1	0	1

Fig 10. Sobel mask used to extract gradient features

Gradient direction in X and Y direction can be calculated as follow by aligning above masks on image I.

$$S_x(x, y) = I(x - 1, y + 1) + 2I(x, y + 1) + I(x + 1, y + 1) \\ - I(x - 1, y - 1) - 2I(x, y - 1) - I(x + 1, y - 1)$$

$$S_y(x, y) = I(x + 1, y + 1) + 2I(x + 1, y) + I(x + 1, y - 1) \\ - I(x - 1, y + 1) - 2I(x - 1, y) - I(x - 1, y - 1)$$

Gradient magnitude is calculated as follow:

$$M(x, y) = \sqrt{S_x^2(x, y) + S_y^2(x, y)}$$

And direction of magnitude at point (x, y) can be found using,

$$\theta(x, y) = \tan^{-1} \left(\frac{S_x(x, y)}{S_y(x, y)} \right)$$

Direction of gradient is decomposed along eight chain code direction and that vector is fed as input to support vector machine.

Velappa et al have proposed multiscale neural network based approach in [15]. Neural networks like Feed forward back propagation neural network requires long training time to memorize and generalize all input feature vectors [10]. And still there are good chances of misclassification. Generalization problem can be overcome by using multi scale neural network [11]. Proposed system first convert camera captured RGB image to binary image. Width to Height Ratio (WH), Relative Height

(RH) ratio, Relative Width ratio (RW) is calculated to remove unnecessary connected components from image. For multi scale neural network, detected character is resized to 20 X 28 pixels, 10 X 14 pixels and 5 X 7 pixels. Binary features of these different resolution images are given to three layer feed forward back propagation algorithm.

In literature [16], T.Som et al have discussed fuzzy membership function based approach for HCR. Character images are normalized to 20 X 10 pixels. Average image (fused image) is formed from 10 images of each character. Bounding box around character is determined by using vertical and horizontal projection of character. After cropping image to bounding box, it is resized to 10 X 10 pixels size. After that, thinning is performed and thinned image is placed in one by one row of 100 X 100 canvas. Similarity score of test image is matched with fusion image and characters are classified.

In literature [17], Rakesh kumar et al has proposed single layer neural network based approach for HCR to reduce training time. Characters are written on A4 size paper in uniform box. Segmented characters are scaled to 80 X 80 pixels. Each 0 is replaced by -1 for better training.

Malayalam is a language spoken by millions of people in the state of Kerala and the union territories of Lakshadweep and Pondicherry in India. It is written mostly in clockwise direction and consists of loops and curves. Neural network based approach is discussed in [11] for Malayalam language. In pre processing step, noise is removed by applying threshold (number of pixels in rectangular bounding box).

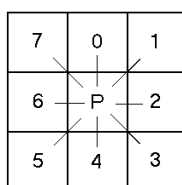


Fig 11. Freeman Code representation

As pen is shifted on paper surface direction is recorded. As shown in fig 5, 1 for North East, 2 for East, 3 for South East etc. These direction values are used as feature vector. Continuous duplicate entries are removed from the feature vector. For improvement, direction is recorded between two points 2 pixel apart, rather than adjacent pixel. It reduces size of feature vector by half. Plus it handles some writing anomalies too. These features are used to train neural network. In post processing, recognized characters are converted to UNICODE format plus disambiguating techniques are applied to distinguish similar characters like 'Pa' and 'Va'.

The complexity of language controls the accuracy of recognition system, whereas this complexity and uncertainty becomes multiplied. The use of Latin languages like English, or Spanish, limits the uncertainty because of the limited structure of the character. Arabic language characters are very complex in comparison with the most languages in the world. Farah et al have discussed Arabic

HCR system in their work mentioned in [18]. Arabic characters are being written continuously connecting to the next or previous character. Segmented character is resized to 48 X 32 pixels block and it is further divided in 6 by 4 segments each of size 8 X 8 as shown in fig 6. Five features are extracted from each of 24 blocks. These 120 features uniquely identify the character.

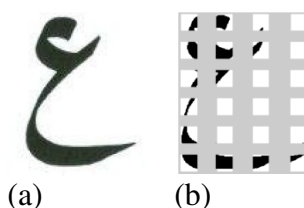


Fig. 12. (a) Original 48 X 32 pixel character, (b). Divided in 24 blocks

These five features are extracted as follows:

- 1). Ratio of white to black pixels
- 2). Ratio of two farthest pixels with respect to diagonal of segment
- 3). Normalized average of spatial segment
- 4). Ratio of Variance of farthest two vertical pixels to variance of farthest two horizontal pixels
- 5). Average value of off diagonal line of segment
- 6). Total variance of segmented pixels

Adapted structure of neural network consists of four separated neural networks; where the first is Multi-Layer Perceptron (MLP) network and the other networks are Linear-Vector Quantizer (LVQ) networks. The linear vector quantizer network has the ability to recognize very close features with lower processing time. In contrast, the multi-layer Perceptron neural network can recognize the unique feature from many different input scopes.

Diagonal based feature extraction work is mentioned in [19], which is improved by Sharma et al, discussed in [20]. They have proposed zone based hybrid feature extraction method. Euler number concept is used to improve speed and accuracy. Thresholding, filtering and thinning operations are performed as a part of preprocessing. Segmentation can be classified into three broad categories: Top down, Bottom up and Hybrid techniques. In proposed method segmented character is resized to 90 X 60. After calculating Euler number from this image, character is divided in to 10 X 10 pixel 54 zones. Each zone value is replaced by average intensity value and is used as feature value. 9 and 6 features are

extracted by averaging values row wise and column wise, so it forms total 69 features. A FFBPNN with configuration 69-100-100-26 is used for classification.

Postal address recognition system for Arabic language is proposed by M.Charfi et al in [21]. Writing translates style of writing, Mood and personality of the writer, which makes it difficult to characterize. From scanned envelop, printed boarder and stamp logo are suppressed. Address is located and using histogram method, lines, words and characters are segmented. Temporal order of strokes can be helpful for robust recognition. In literature, way of temporal order reconstruction is proposed. End stroke point, Branching point and Crossing point are detected from city name. Elliptical model is applied on preprocessed digit or character and matching process is applied.

In literature [22], RenataF. P. Neves et al have proposed SVM based offline handwritten digit recognition. Authors claim that SVM outperforms the Multilayer perceptron classifier. Experiment is carried out on NIST SD19 standard dataset. Advantage of MLP is that it is able to segment non-linearly separable classes. However, MLP can easily fall into a region of local minimum, where the training will stop assuming it has achieved an optimal point in the error surface. Another hindrance is defining the best network architecture to solve the problem, considering the number of layers and the number of perceptron in each hidden layer. Because of these disadvantages, a digit recognizer using the MLP structure may not produce the desired low error rate

Amritha et al have discussed freeman code based Malayalam online handwritten character recognition in their literature [23]. It is also known as chain code. Chain code is widely used features for character recognition. Chain codes are the directional features. Starting from the point when first contact is made with the writing surface, direction in which the pen tip moves is recorded. 1 for NE, 2 for E, 3 for SE etc. will be stored as a single dimensional array. Direction is recorded only when there is change in direction to avoid dependence of length of line segments in the character. This array is used as a feature vector for classification. An issue, that arise when creating the feature vector based on direction of pen movement is that, instead of storing a '1' in the feature vector for the NE direction, it may store it as '2' followed by '0'. This issue arises due to irregularities in writing caused due to the inexperience of the user in using the device, shivering during writing etc. This can be avoided by extracting the direction formed between points 2 pixels apart rather than adjacent pixels. This greatly helps to reduce the size of feature vector and makes it more accurate.

In literature [24], [25], pritpalsingh et al has mentioned wavelet transform based handwritten character and numeral recognition for Gurumukhi script. Color images are converted in gray scale and median filter is applied to remove noise. Binarized image is then normalized to 32 X 64 pixels size. Wavelets are localized basis functions which are translated and dilated versions of some fixed mother wavelet. The decomposition of the image into different frequency bands is obtained by successive low-pass and high-pass filtering of the signal and down-sampling the coefficients after each filtering. In this

work, various Discrete Wavelet Transforms e.g. db1, db4, sym2, sym4, coif3, coif5, bior1.3, and bior3.9 have been used to extract the wavelet coefficients.

Dileep et al have proposed multi resolution technique using Discrete Wavelet Transform (DWT) based approach for HCR in their work [26]. Representation of images in various degrees of resolution is known as multi resolution process. Wavelets are used as the foundation of multi resolution process. Scanned handwritten characters are normalized to 100 X 100 pixels. Binary image is thinned using skeletonization infinite times. Edges of these thinned images are detected using appropriate thresholding and then further dilated using appropriate structure element. Multi resolution technique is applied on this diluted Image. DWT has excellent spatial localization and good frequency localization properties that makes it an efficient tool for image analysis. Many multi resolution techniques like image pyramids, sub band coding, DWT etc. are available. Authors have used DWT based multi resolution technique, which maps continuous variable in to sequence of coefficients. Below figure explains the preprocessing stage of proposed method.

Muhammad et al have discussed hybrid feature extraction in their work mentioned in [27]. SVM is used as a classifier. Authors have combined structural, statistical and correlation functions to derive hybrid features. In first step, elementary stroke location is identified with the help of chosen elementary shape. To make it more robust, certain structural / statistical features are added in it. The added structural / statistical features are based on projections, profiles, invariant moments, endpoints and junction points. This enhanced, powerful combination of features results in a 157-variable feature vector for each character. It includes 100 correlation features and 57 structural/statistical features. Correlation features are based on Pearson's correlation coefficient.

Shubhangi et al [28], have extract similar correlation function based features for Chinese hand-printed character recognition. Classification is done based on minimum distance decision rule. While proposed method perform final classification based on support vector machine (SVM).

TABLE I: COMPARISION OF VARIOUS TECHNIQUES FOR HCR

Paper	Language	Online	Offline	Model / Features	Dataset Size	Accuracy Claimed
[8]	English	-	√	HMM Gradient (4), Projection (6), Curvature (6), Local Grad (4)	100 Writers X 5 sample of each = 13000 Samples (26 Alphabet) Training Set : 2600	98.26 %

					Char Testing Set: 10,400	
[11]	Malayalam	√	-	FFBPNN Freeman Code		
[14]	Hindi			Gradient SVM	20 Writer X 10 Samples = 7200 Samples (36 Alphabets) Normalized Size: 90 X 90	94.00 %
[15]	English	-	√	Multi scale Neural Network	10 sample X 26 Alphabets = 260 10 X 10 numeric = 100	85.00 %
[16]	English		√	Fuzzy Membership based function	10 sample X 26 char = 260	
[17]	English			Single Layer NN. Binary features	26 Characters only	80.00 %
[18]	Arabic	-	√	Modified Cascaded ANN 5 features from each block of character	Total 100 Characters from 10 persons	67.00 %
[20]	English		√	Zone based hybrid feature extraction model		

[21]	Arabic		$\sqrt{}$	Beta Elliptical Model		98.00 %
------	--------	--	----------------------	--------------------------	--	---------

IV. DISCUSSION AND CONCLUSIONS

Comparison between all discussed literatures is shown in table I with all possible dimensions. Literature review shows that neural network is the prime choice among researchers for training purpose. However various kind of changes have been proposed in feature extraction methods. Work can be extended from single character or set of characters to document processing. Many applications are awaiting the enhancement in character recognition to be adopted it fully. Hybrid model can be proposed which counts on more than one feature extraction methods to discriminate characters properly. Post processing is also crucial to discriminate similar structured characters like Q and O, M and N like pairs in English language.

It is not possible to design fully automated system which can handle all kind of variability. HCR system must be domain specific as well as bounded to limited versatility in writing style for robustness. Handwritten character recognition is still a burning research area of pattern recognition. Each and every step contributes directly to the accuracy of system, like preprocessing, segmentation, feature extraction, training methods etc all. So all these area are open for independent research. A lot can be improved in each of step.

REFERENCES

- [1] Kai Ding, Zhibin Liu, Lianwen Jin, Xinghua Zhu, "A Comparative study of GABOR feature and gradient feature for handwritten Chinese character recognition", International Conference on Wavelet Analysis and Pattern Recognition, pp. 1182-1186, Beijing, China, 2-4 Nov. 2007
- [2] Pranob K Charles, V.Harish, M.Swathi, CH. Deepthi, "A Review on the Various Techniques used for Optical Character Recognition", International Journal of Engineering Research and Applications, Vol. 2, Issue 1, pp. 659-662, Jan-Feb 2012
- [3] Om Prakash Sharma, M. K. Ghose, Krishna Bikram Shah, "An Improved Zone Based Hybrid Feature Extraction Model for Handwritten Alphabets Recognition Using Euler Number", International Journal of Soft Computing and Engineering, Vol.2, Issue 2, pp. 504-58, May 2012
- [4] J. Pradeepa, E. Srinivasana, S. Himavathib, "Neural Network Based Recognition System Integrating Feature Extraction and Classification for English Handwritten", International journal of Engineering, Vol.25, No. 2, pp. 99-106, May 2012
- [5] Liu Cheng-Lin, Nakashima, Kazuki, H, Sako, H.Fujisawa, "Handwritten digit recognition: investigation of normalization and feature extraction techniques", Pattern Recognition, Vol. 37, No. 2, pp. 265-279, 2004.
- [6] Supriya Deshmukh, Leena Ragha, "Analysis of Directional Features - Stroke and Contour

- for Handwritten Character Recognition", IEEE International Advance Computing Conference, pp.1114-1118, 6-7 March, 2009, India
- [7] Amritha Sampath, Tripti C, Govindaru V, "Freeman code based online handwritten character recognition for Malayalam using Back propagation neural networks", Advance computing: An international journal, Vol. 3, No. 4, pp. 51-58, July 2012
- [8] Rajib Lochan Das, Binod Kumar Prasad, Goutam Sanyal, "HMM based Offline Handwritten Writer Independent English Character Recognition using Global and Local Feature Extraction", International Journal of Computer Applications (0975 – 8887), Volume 46– No.10, pp. 45-50, May 2012
- [9] Bhatia, N. and Vandana, "Survey of Nearest Neighbor Techniques," International Journal of Computer Science and Information Security, Vol. 8, No. 2, (2001), 302-305.
- [10] Rajbala Tokas, Aruna Bhadu, "A comparative analysis of feature extraction techniques for handwritten character recognition", International Journal of Advanced Technology & Engineering Research, Volume 2, Issue 4, pp. 215-219, July 2012
- [11] Amritha Sampath, Tripti C, Govindaru V, "Freeman code based online handwritten character recognition for Malayalam using backpropagation neural networks", International journal on Advanced computing, Vol. 3, No. 4, pp. 51 - 58, July 2012
- [12] Tirtharaj Dash, "Time efficient approach to offline hand written character recognition using associative memory net.", International Journal of Computing and Business Research, Volume 3 Issue 3 September 2012
- [13] Imaran Khan Pathan, Abdulbari Ahmed Bari Ahmed Ali, Ramteke R.J., "Recognition of offline handwritten isolated Urdu character ", International Journal on Advances in Computational Research, Vol. 4, Issue 1, pp. 117-121, 2012
- [14] Ashutosh Aggarwal, Rajneesh Rani, RenuDhir, "Handwritten Devanagari Character Recognition Using Gradient Features", International Journal of Advanced Research in Computer Science and Software Engineering (ISSN: 2277-128X), Vol. 2, Issue 5, pp. 85-90, May 2012
- [15] Velappa Ganapathy, Kok Leong Liew, "Handwritten Character Recognition Using Multi scale Neural Network Training Technique", World Academy of Science, Engineering and Technology, pp. 32-37, 2008
- [16] T.Som, Sumit Saha, "Handwritten Character Recognition Using Fuzzy Membership Function", International Journal of Emerging Technologies in Sciences and Engineering, Vol.5, No.2, pp. 11-15, Dec 2011
- [17] Rakesh Kumar Mandal, N R Manna, "Hand Written English Character Recognition using Row- wise Segmentation Technique", International Symposium on Devices MEMS, Intelligent Systems & Communication, pp. 5-9, 2011.
- [18] Farah Hanna Zawaideh, "Arabic Hand Written Character Recognition Using Modified

- Multi-Neural Network", Journal of Emerging Trends in Computing and Information Sciences (ISSN 2079-8407), Vol. 3, No. 7, pp. 1021-1026, July 2012
- [19] J Pradeep, E Shrinivasan and S.Himavathi, "Diagonal Based Feature Extraction for Handwritten Alphabets Recognition System Using Neural Network", International Journal of Computer Science & Information Technology (IJCSIT), vol. 3, No 1, Feb 2011.
- [20] Om Prakash Sharma, M. K. Ghose, Krishna Bikram Shah, "An Improved Zone Based Hybrid Feature Extraction Model for Handwritten Alphabets Recognition Using Euler Number", International Journal of Soft Computing and Engineering (ISSN: 2231 - 2307), Vol. 2, Issue 2, pp. 504-508, May 2012
- [21] Moncef Charfi, Monji Kherallah, Abdelkarim El Baati, Adel M. Alimi, "A New Approach for Arabic Handwritten Postal Addresses Recognition", International Journal of Advanced Computer Science and Applications, Vol. 3, No. 3, pp. 1-7, 2012
- [22] Renata F. P. Neves, Alberto N. G. Lopes Filho, Carlos A.B.Mello, CleberZanchettin, "A SVM Based Off-Line Handwritten Digit Recognizer", International conference on Systems, Man and Cybernetics, IEEE Xplore, pp. 510-515, 9-12 Oct, 2011, Brazil
- [23] Amritha Sampath, Tripti C, Govindaru V, "Freeman code based online handwritten character recognition for Malayalam using Back propagation neural networks", Advance computing: An international journal, Vol. 3, No. 4, pp. 51-58, July 2012
- [24] Pritpal Singh, Sumit Budhiraja, "Handwritten Gurumukhi Character Recognition using Wavelet Transforms", International Journal of Electronics, Communication & Instrumentation Engineering Research and Development, Vol. 2, Issue 3, pp. 27-37, Sept 2012
- [25] Pritpal Singh, Sumit Budhiraja, "Offline Handwritten Gurmukhi Numeral Recognition using Wavelet Transforms", International Journal of Modern Education and Computer Science, vol. 8, pp. 34-39, August 2012
- [26] Dileep Kumar Patel, Tanmoy Som, Sushil Kumar Yadav, Manoj Kumar Singh, "Handwritten Character Recognition Using Multiresolution Technique and Euclidean Distance Metric", Journal of Signal and Information Processing, Vol. 3, pp. 208-214, May 2012
- [27] Muhammad Naeem Ayyaz, Imran Javed, Waqar Mahmood, "Handwritten Character Recognition Using Multiclass SVM Classification with Hybrid Feature Extraction", Pakistan journal of Engineering and Application Science, Vol. 10, pp. 57-67, Jan – 2012
- [28] D. C. Shubhangi and P. S. Hiremath, "Handwritten English character and digit recognition using multiclass SVM classifier and using structural micro features," International Journal of Recent Trends in Engineering, vol. 2, no. 2, Nov. 2009.