



# A KNN Improved Art Network Approach for Handwritten Character Recognition under Noise

<sup>1</sup>Neha, <sup>2</sup>Anil Saroliya, <sup>3</sup>Varun Sharma

<sup>1, 2, 3</sup>Amity School of Engineering & Technology, Amity University Rajasthan, Jaipur, INDIA

*Abstract- One of the major application of Image processing is the character recognition. Character recognition is effective to convert the image character to text. This recognition process is more challenging in case of noisy or disturb image. In this work, a hybrid approach is suggested to perform this recognition effectively in case of disturbed noise image. The work is divided in three main stages. In first stage, the image improvement is performed by using denoising algorithm. In second stage, the image feature extraction is done to identify the character ROI and the feature points. This feature extraction is defined using KNN approach. At the final stage, Art network is applied to perform the image classification and recognition under effective vigilance ratio. The obtained results from system show the effective recognition rate.*

*Keywords: OCR, KNN, Art Network, Feature Extraction*

## I. INTRODUCTION

Digital Character Recognition is having its importance in many of the real time applications. In many of the business applications, the information such as cost, brand is presented in printed form.

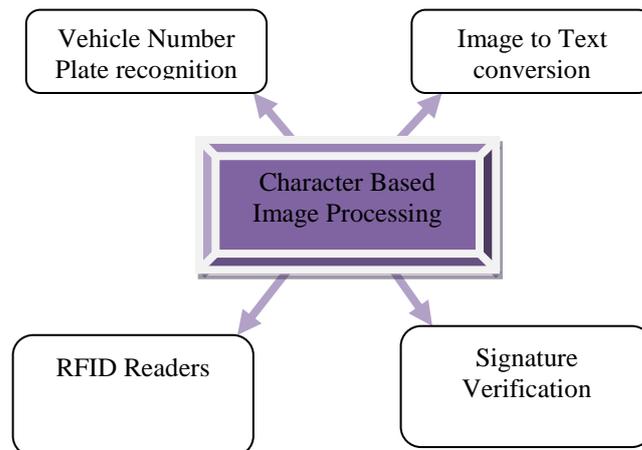


Figure 1 : Character Processing Applications

To capture and read this information in automated way, there is the requirement of some such mechanism that can capture and read the image and convert it to textual form effectively. Some of the common character processing applications are shown in figure 1. These all applications are having its significance to boost up the process by providing the automated data reading. But while performing such kind of recognition, one of the common image processing problems is the bad image identification. This bad image can be a noisy or disturbed image that is does not explore the image features effectively. The reason of such kind of capturing includes the bad situation, light or the camera. These kind of problem includes the speckles or noisy while performing the image capturing. The presented work is about to define an effective hybrid approach to improve the character recognition process.

The main utility of character recognition system is to classify the digital and optical patterns so that the alphanumeric character will be obtained. To perform this recognition a series of operations are adapted such as segmentation, feature extraction and classification. Based on these operations the actual recognition of character is performed. The scanning is also been under the human recognized characters and text so that the effective detection of the character will be performed.

OCR is one of the traditional fields of pattern recognition, artificial intelligence and the machine vision. In case of academic system and library management, the significance of the OCR recognition is proven already. The character recognition is basically performed using the mirrors or the lenses. The character recognition is considered as the separate field so that the recognition of characters will be done effectively. OCR is defined as the important image processing application in which the recognition is based on multiple parameters. These parameters include the feature extraction and feature specification. The features depend on the algorithmic approaches adapted to extract the features.

In such kind of systems, the recognition accuracy is one of the important things. Here the input is passed as the raw data and the recognition is based on the scanned image so that the effective textual components will be extracted from the image. This recognition system deals with the character identification and the scanning operation. The document is processed for feature extraction and recognition. The final outcome is derived in the form of a picture box containing the recognized character. The challenges associated with these kind of recognition system is shown in figure 2.

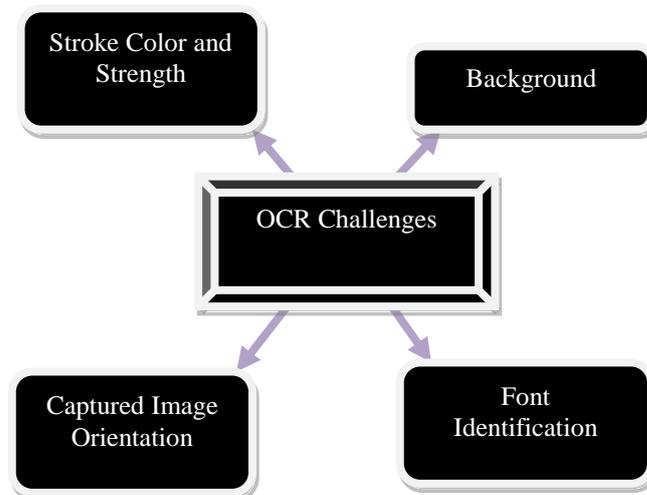


Figure 2 : Associated Challenges

In this paper, an effective hybrid approach is presented to perform the digital character recognition for noisy image. This recognition system is defined using KNN improved Art Network Approach. In this section, introduction to OCR processing is defined along with application and challenge specification. In section II, the work defined by earlier researchers is discussed. In section III, the proposed research work is defined. In section IV, the results obtained from the work are discussed.

## II. RELATED WORK

Image processing is having its valuable importance in the OCR based applications. These applications include the character identification and classification. The work already done by different researchers in this area is discussed in this section.

David Andre[1] presented a work on rule updation based on learning approach in OCR system using Genetic approach. Author defined the genetic programming approach for effective character identification. Author defined a human hand coded rules for initial population generation and rule updation. Author analyze the work on different datasets under different real time problems such as noise etc. Der-Sheng Lin[2] presented a work using genetic algorithm to perform handwritten Chinese character recognition. The work includes he normalized transformation to assign the weighted. Based on this weighted analysis, effective recognition over the digital characters is performed. Angelo Marcelli[3] presented a work based on structural analysis to perform the shape based recognition to apply the effective encoding and transformation so that the effective vector space will be generated and processed under genetic approach. The vector based structural analysis is performed under genetic approach to perform the recognition. Tim J. Klassen[4] has presented an effective recognition process for Arabic characters. Author defined the work for online and offline character recognition. Author presented the SOM based heuristic approach to perform feature anlaysis on online data so that the effective recognition will be obtained. Author presented the genetic based approach to improve the recognition process. Kejian Wang[5] has defined research on Chinese character recognition using genetic approach. Author presented the effective character recognition in real environment.

D. Impedovo[6] has presented a survey on different segmentation approaches to recognize hand written characters. This work includes the analytical study of different zoning methods respective to topological and membership approaches. These approaches are divided under different categories such as static and dynamic approaches, uniform and non-uniform approaches. The work also includes the parameter based analysis on different member ship approaches to different zones. To decide this membership, the approaches considered in this work includes the border based approaches, ranked approaches, fuzzy rule based analysis etc. These approaches include the feature based analysis including the local as well as global features. Author defined the comparative analysis over these approaches and suggested the effective approach under the analytical exploration. Another work on the handwritten character recognition was defined by Depeng[7]. This work includes is performed using the locality alignment approach for discriminative Chinese characters. The work is performed using a hybrid learning approach performed using locality alignment and subspace analysis. The defined approach was kernel based approach that used the PCA as the initial stage and followed by discriminative locality analysis approach. The obtained results shows that the proposed approach had provided the effective accuracy. Another work on the feature extraction based handwritten character recognition was defined by Amit Choudhary[8]. This work includes the feature factor analysis under the binarization approach to improve the accuracy of the recognition process. Once the features are extracted, a multi layered neural network approach is implemented as the character classification and recognition. Author obtained the promising results from the defined system.

Jia Zeng[9] has defined a character modelling and recognition approach based on the statistical structured analysis and markov model based recognition process. This structured analysis includes the stroke analysis, neighbourhood character analysis and the encoding technique so that effective recognition will be performed. The recognition process is performed on the certain features using the markov model based predictive approach. The work is implemented on Korean dataset and the obtained results shows that the work has improved the accuracy of the system. Another comparative study on different recognition and classification approaches for handwritten devnagri character recognition was done by U. Pal[10]. In this work author presented analysis on 12 classifiers with 4 feature sets. These feature sets includes the projection distance analysis, subspace method, linear discriminant function etc. Author performed the analytical study under different information analysis such as survature based and the gradient information analysis. Yuefeng Chen[11] has defined an artificial immune system based handwritten character recognition. Author defined the analysis over the optimization of rate and time for the recognition. This approach is based on the biological principle with the memory cell based analysis. Author presented the experimentation on UCI dataset. The adaptive algorithm provided by the author had improved the speed and accuracy. It was able to identify the pattern recognition and abnormality detection. Another wok on the dictionary based analysis to perform character recognition was done by Shinji Tsuruoka[12]. Author defined a separate library set for each author to identify the writing similarity. Author defined the character specific analysis along with feature space generation so that effective covariance matrix will be generated. Author defined the work on Japanese characters so that effective recognition will be done.

## III. Proposed Approach

In this section, the proposed hybrid model is presented to perform the recognition. As the earlier stage, the training set is defined on which the feature extraction is performed and the featured dataset is generated. Once the feature dataset is obtained, the noisy input image is captured to perform the recognition process. To perform this recognition, the input image is pre-processed. This pre-processing stage includes the denoising algorithm and the character area identification over the image. To remove the image

noise, Gaussian filter is applied in this work and to perform the image segmentation, the combination of mathematical filters is applied. These mathematical filters include the convolution filter and morphological filters. Based on these filters, the character area from the image is extracted. At the final stage, the KNN improved Art network approach is applied to perform the recognition and classification. The KNN is here used for feature map and based on this mapping, vigilance vector of the recognition process is obtained. Now the vigilance ratio match is performed using art network to identify the character class based on the dataset classes. The effective matching class is identified as the recognised character image. The basic model of this presented work is shown in figure 3.

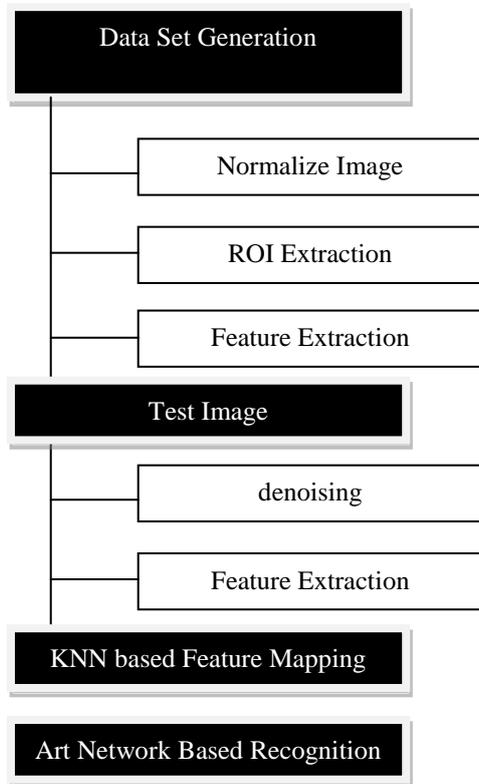


Figure 3 : Character Image Recognition

The basic algorithm approaches used in this work are given here under. The complete work is divided in two main algorithmic stages. The Gaussian filter based denoising and hybrid recognition algorithm. These algorithmic approaches are defined in this section.

**A) Gaussian Filter**

As the presented work is defined to perform the recognition under noisy input image. The denoising is here performed using Gaussian filter. The algorithmic approach for Gaussian filter is shown in figure 4. The Gaussian filter is more robust compared to the mean and median filter. Thus, a single very unrepresentative pixel in a neighborhood will not affect the median value significantly. Since the FFT based decoding process is defined to obtain the signal values. This value must actually be the value of one of the pixels in the neighborhood; the filter does not create new unrealistic pixel values when the filter straddles an edge. For this reason the median filter is much better at preserving sharp edges than the mean filter. These advantages aid median filters in denoising uniform noise as well from an image. The denoising approach is effective for additive noise as well as on multiplicative noise. The Flowchart of the work is shown here under.

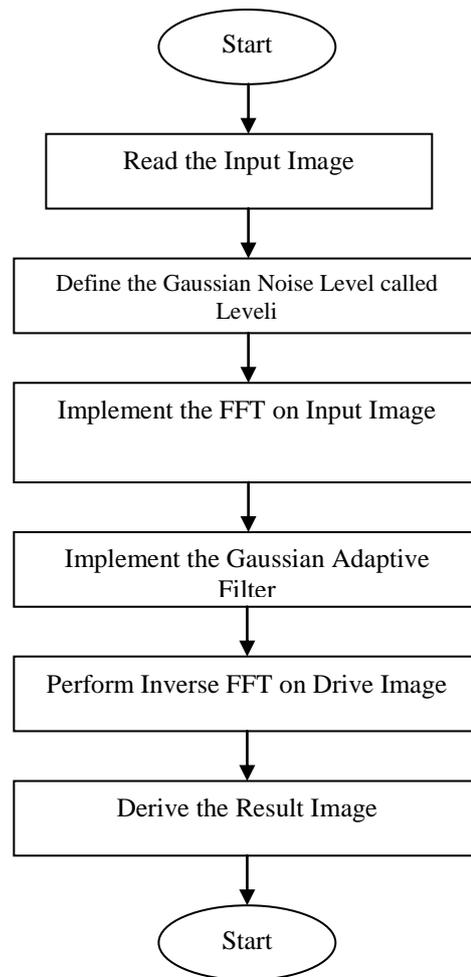


Figure 4 : Gaussian Filter

**B) Recognition**

The recognition is here defined using KNN based ART network. The recognition is here defined as the feature based vigilance match of input image with dataset images. The dataset is trained at initial stage and the vigilance values are obtained and the vigilance value dataset is generated. The algorithmic approach for recognition process is shown in table 1

Table 1 : Recognition Algorithm

1.	We have a Trained Art Network with N Classes
2.	Input Image Img
3.	Define Vegilience Vector V
4.	matchratio=0;
5.	p=null /* initialize the match image*/
6.	for c=1 to N
7.	{
8.	img1=GetImage(c)
9.	Find Feature Difference Diff=img1-img
10.	M=Matchingratio(img,img1)

```

11.   if Differecne>=V and M<matchratio
12.   {
13.   matchratio=M;
14.   p=img1;
15.   }
16.   }

17.   if(p==null)
18.   {
19.   Print “No Match Image Found”
20.   }
21.   else
22.   {
23.   Print “Image Detected “+ p
24.   }
25.   }
    
```

#### IV. RESULTS

The presented work is applied on digital alphanumeric characters defined in gray scale. The properties of the dataset are shown in table 2.

Table 2: Dataset Properties

Parameter	Value
Number of Images	26
Color	No
Images Type	Alphabet
Image Size	100x100
Image Format	BMP
Image Fault/Noise	Yes ( Level .1)
Image Filtration	Gaussian
Recognition	KNN Improved Art Network

The recognition process is here defined to perform the recognition. The recognition is here defined at class level under vigilance vector so that effective recognition will be performed. The recognition property set is shown here under.

Table 3 : Recognition Properties

Properties	Values
Number of training Images	26
Number of Test Images	12
Noisy Images	5
Correctly Detected	11
Noisy Correctly Detected	4
Non Noisy Correctly Detected	7
Recognition Rate Noisy Images	100%
Recognition rate (Noisy Images)	80%

## V. CONCLUSION

In this paper, a KNN improved ART network approach is defined to perform the recognition. The work is here defined for English alphanumeric characters. The work is effective for noisy images. The recognition rate obtained from the work shows the effective detection of objects.

## References

- [1] David Andre," Learning and Upgrading Rules for an OCR System Using Genetic Programming", 0-7803-1899-4194@1994 IEEE
- [2] Der-Sheng Lin," A Genetic Algorithm Approach to Chinese Handwriting Normalization", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 27, NO. 6, DECEMBER 1997 1083–4419/97 ã 1997 IEEE
- [3] Angelo Marcelli," Exploring genetic programming for modeling character shape", 0-7803-6583-6/00@2000 IEEE
- [4] Tim J. Klassen," Towards the On-line Recognition of Arabic Characters", 0-7803-7278-6/02@2002 IEEE
- [5] KEJIAN WANG," RESEARCH ON CHINESE CHARACTER RECOGNITION POSTPROCESSING BASED ON GENETIC ALGORITHM", Proceedings of the First International Conference on Machine Learning and Cybernecics, Beijing, 4-5 November 2002 0-7803-7508-4/02@2002 IEEE
- [6] D. Impedovo, Zoning methods for handwritten character recognition: A survey, PatternRecognition, pp 969–981, 2014
- [7] Dapeng Tao, Similar handwritten Chinese character recognition by kernel discriminative locality alignment, Pattern Recognition Letters, pp 186–194, 2014
- [8] Amit Choudhary, Off-Line Handwritten Character Recognition using Features Extracted from Binarization Technique, AASRI Conference on Intelligent Systems and Control, pp 306 – 312, 2013
- [9] Jia Zeng, Markov Random Field-Based Statistical Character Structure Modeling for Handwritten Chinese Character Recognition, IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 30, No. 5, pp 767-780, 2008
- [10] U. Pal, Comparative Study of Devnagari Handwritten Character Recognition using Different Feature and Classifiers, 10th International Conference on Document Analysis and Recognition, pp 1112-1115, 2009
- [11] Yuefeng Chen, A Handwritten Character Recognition Algorithm based on Artificial Immune, International Conference on Computer Application and System Modeling, vol 12, pp 273-276, 2010
- [12] Shinji Tsuruoka, Personal Dictionaries for Handwritten Character Recognition Using Characters Written by a Similar Writer, 12th International Conference on Frontiers in Handwriting Recognition, pp 599-604, 2010
- [13] Soumen Bag, Recognition of Bengali Handwritten Characters Using Skeletal Convexity and Dynamic Programming, Second International Conference on Emerging Applications of Information Technology, pp 265-268, 2011