ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

Offline Handwritten Numeral Recognition using Combinational Feature Extraction Approach

Nishu Rastogi[1], Maitreyee Dutta[2], Ajay Indian[3]

Assistant Professor, Department of Computer Applications, Invertis University, Bareilly, India[1]

Professor and Head, Department of Electronics and Communication Engineering, NITTTR, Chandigarh, India[2]

Associate Professor, Department of Computer Applications, Invertis University, Bareilly, India[3]

ABSTRACT: Handwritten Character Recognition has become a very challenging area of research, the challenge is introduced just because of the fact that the handwriting of a person differs very much from another person, even the handwriting of a person varies depending upon the mood and environment. The conclusion drawn from existing work is that recognition accuracy depends on how and what features are utilized to formulate the feature vector from the sample set. A combinational feature vector contributes towards increased recognition rate. In this paper, we have proposed and implemented combinational approach of feature extraction which includes direction gradient as well as 8 – directional Quadrant Based Chain Code Histogram (QBCCH) to elicit structural components from the samples and the results are calculated by formulating feature vector individually as well as with their combinations. For training and testing purpose, samples of Gujarati numerals from 0 to 9 are considered as the pattern of them is a blend of both Hindi and English numerals. A feature vector of length nine using directional gradient histogram (DGH), eight using chain coding (CCH) and thirty two using chain code by dividing image into four parts (QBCCH) is passed through the Back-propagation Neural Network with Levenberg-Marquartdt training function and recognition rate of approx. 72%, 82% and 98% respectively is attained. With combinational feature vector using DGH with CCH and DGH with QBCCH resulted in much higher recognition rates for isolated Gujarati Numerals.

KEYWORDS: Handwritten Character Recognition**;** Gradient**;** Freeman's Chain Code**;** Back-propagation Neural Network.

I. INTRODUCTION

Character Recognition means to detect or identify either the printed or a handwritten character. Character set includes all the numeric values i.e. from 0 to 9 and alphabets as well as special symbols. Recognizing the printed character is still an easier task as the font styles and families are limited while in case of handwritten text recognition becomes strenuous as there is no restriction on the styles of writing the text by a human being. Other than this distinction text may be recognized during the process of writing or afterwards. If the process of recognition is applied during the time text is being written is known as On-line Character Recognition otherwise if recognition process is applied after the text has been written is known as Off-line Character Recognition [12]. The process of character recognition involves several sub-routines that is sample collection from different people of different age group, applying pre-processing on the collected samples to make them ready to extract the features, an effective and efficient algorithm is needed to be applied to extract feature and for feature vector creation, once the feature vector is created now is the turn to make use of some classification algorithm to assign particular character to a specific class. A number of feature extraction techniques have been proposed by the researchers to mine the specific features. Feature extraction methods for handwritten character recognition are based on either structural features or statistical features. [13][14][15] For classification and training purpose several algorithms are available like BPNN, CNN, SVM, and Genetic Algorithms etc. Fig. 1 represents various steps which are to be followed for character recognition.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

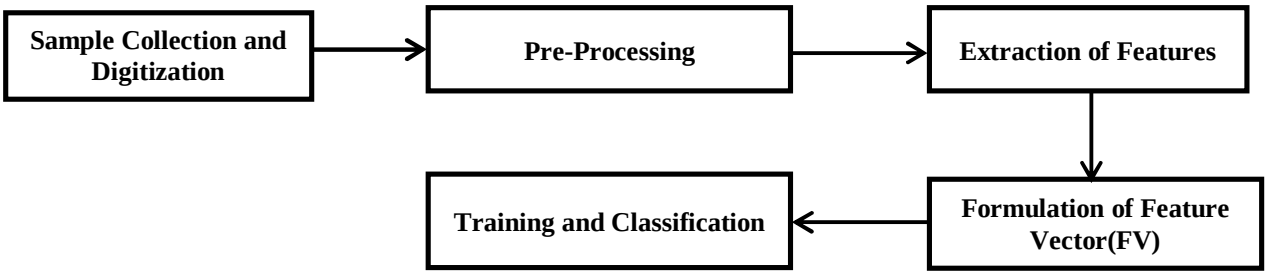


Fig.1 A Character Recognition System

II. RELATED WORK

A number of researchers are working in this field of character recognition which is a part of pattern recognition. The main aim of the researchers is to improve the recognition rate. Use of an efficient algorithm for feature extraction and a reduced feature vector can result in less complexity and improved accuracy in classification. As character recognition involves several steps and at each level a new innovation can be done to raise the recognition rate. [1] Dayashankar et al. have proposed a scheme that calculates the directional gradient feature and normalized the values to eight directions and resulted in recognition rate of 90%. The feature vector (i.e. length 900) was very large in this works which needs high computing capabilities. [2]AshutoshAggarwal et al. have reduced the length of feature vector by considering 8 directional gradients [16] by normalizing and diving image into small blocks, so reduced length of FV from 900 to 200 and obtained the rate approx. 97% using the SVM classifier with RBF kernel for Gurumukhi characters. [3] Feature vector is created utilizing both statistical and structural component from the sample set. They included image centroid, run length code, aspect ratio and gradient features to produce the feature vector of length 147. Recognition rate of 99% on Malayalam characters was achieved. It shows that using a combinational feature vector may result in increased accuracy. [4] Xianjing et al. have proposed a circular grid zoning method on Polar transformation and achieved 92.3% accuracy. [5] Shen-Wei lee et al. have used combinational technique for feature extraction and noticed recognition rate of 98% through SVM classifier. As per H. Imran et al. in [6] the 4 directional / 8 directional chain code is a better way to extract the structural feature and for training purpose K-NN classification model was used, a very high recognition rate was observed by them. In [7] by Qian You et al. have implemented chain code histogram feature and a differential chain code histogram feature by judging the main turning points from the direction movement of character. Feature vector is passed through SVM. In [8] Jomy John et al. have implemented the chain code scheme on Malayalam character and observed approx. 61 % and 63 % of accuracy using 4-neighborhood chain code respectively. So to improve accuracy they have included on more feature with chain code i.e. the centroid and attained the recognition rate of approx. 66% and 72 % with 4and 8 neighborhood normalized chain code. Here we have considered Gujarati Numerals due to their similar structure and it has been noticed that a very less work has been done on Gujarati and recognition arte is also very less for it [9-11],[17-18].

III. PROPOSED METHODOLOGY

The main objective of the research work is to find an efficient algorithm for feature extraction with reduced the length of feature vector as well as which can increase the accuracy of recognition of Gujarati Numerals.
For us it is very easy to recognize printed as well as handwritten character as we as do not recognize things on the basis of exact structure or pattern of any entity. We have a capability to derive the common patterns very quickly through the Intelligence gifted to us by God. But to write such an intelligent program for computer to show such characteristic like humans is a difficult and a tedious job because computer work on comparison if the two entities are exactly the same it can identify otherwise it will not. To make the computer intelligence it is necessary to first analyze and detect the

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

procedure following which human gets able to remind or recall or identify things i.e. the uniqueness among them which is referred as feature vector. Then the next task is to make the computer learn it to get able to make judgments by adjusting the values.

Proposed Algorithm and flowchart is as follows-

Step 1- Collect a number of samples of handwritten Gujarati Numerals from different people of different age group as well as from left handed and right handed persons.
Step 2- Digitize all the samples by scanning them.
Step 3- Convert the image into binary image (40 x 40 pixels).
Step 4- Segment the image using bounding box and isolate numeral.
Step 5- Calculate Chain code and Gradient direction.

A) For Chain Code of character

i- Apply morphing operation on the image to extract only the boundary of character.
ii- Perform padding on the resultant to separate edge of character and boundary of image.
iii- Compute the chain code (i.e. 8-neighborhood, direction movements during writing with respect to previous pixel also known as Freeman's Chain code).
iv- Normalize the chain code i.e. Chain Code Histogram(CCH) and Quadrant Based Chain Code Histogram i.e. QBCCH (by storing the frequency of each direction for whole image and quadrant-wise also)

Example 0 – 0 – 0 – 7 – 7 – 7 – 4 – 4 – 4 – 2 – 2

Count of 0=3; 1=0; 2=2; 3=0; 4=3; 5=0; 6=0; 7=3.

v- A feature vector will be obtained for complete image and quadrant-wise of length 8 and 32 (i.e. 8 x4 = 32) respectively.

B) **For Gradient Direction of character**

i- Calculate Gradient direction using Sobel operator.
ii- Round off the directional gradient.
iii- Normalize the direction to get Directional Gradient Histogram i.e. DGH, consider the class for directions as follows-

Table.1 Directional Gradient Classes

Gdir	Class	Gdir	Class
45 > = Gdir > 0	1	270 >= Gdir > 225	6
90 > = Gdir > 45	2	315 >= Gdir > 270	7
135 > = Gdir > 90	3	360 >= Gdir > 315	8
180 >= Gdir > 135	4	Gdir = 0	0
225 >= Gdir > 180	5		

iv- A feature vector of length nine will be obtained.

Step 6- After extracting the features from various sample, create a feature vector including both directional group and chain code.
Method A- Combine Feature vector of chain code of length 8 with gradient direction of length 9 which will result in a feature vector of length 8 + 9 = 17.
Method B- Combine Feature vector of chain code calculated for quadrant of length 32 with gradient direction of length 9 resulting in a feature vector of 32 + 9 = 41
Step 7- Pass the features vector created using Method A and Method B through the Feed Forward Back-propagation Neural Network for training.
Step 8- Compare the recognition rate by inputting a new image to the trained networks to test which gives good accuracy.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

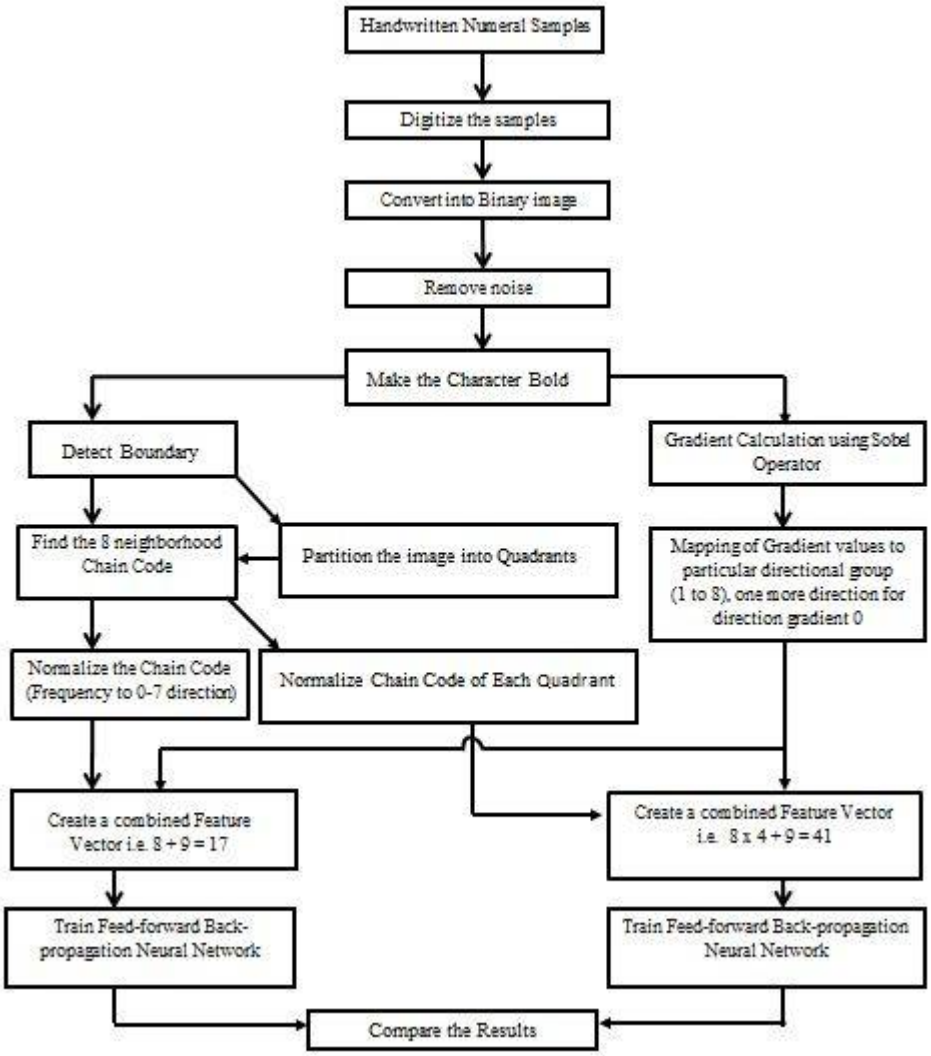


Fig.2 Flowchart of Proposed Methodology

IV. IMPLEMENTATION AND RESULTS

The samples of 0-9 Gujarati numerals were collected on the piece of paper from people of different age group. Experiment was performed on a sample set of 2500 containing approx. 250 sample of each numeral. All the samples taken on the paper were converted to digital format (.jpg) by scanning the sheets through a flatbed Cannon scanner with an optical resolution of 2400 x 4800 dpi. Gujarati numerals sample is considered here due to their complex and resemblance with Hindi and English numeral. Fig. 2 shows the sample set of Gujarati numeral and Fig. 3 shows a sample collected and used for experiment purpose.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

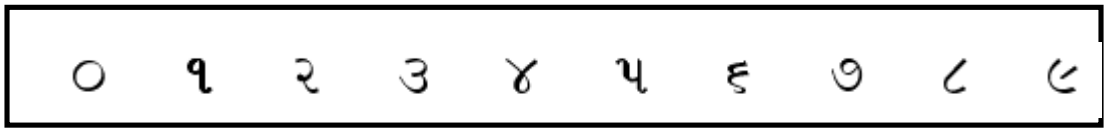


Fig. 2 A Sample Set of Gujarati Numeral

After digitizing the samples preprocessing is done to highlight the features of character. Features vector has been derived by applying the algorithm as discussed above. For training purpose two hidden layer Feed-forward Back-propagation Neural Network with Levenberg-Marquartdt function. Neural Network Once the neural network is trained it is ready to predict and recognize the new handwritten numeral of Gujarati. Neural networks were trained using the individual methods as well as using the compound features.

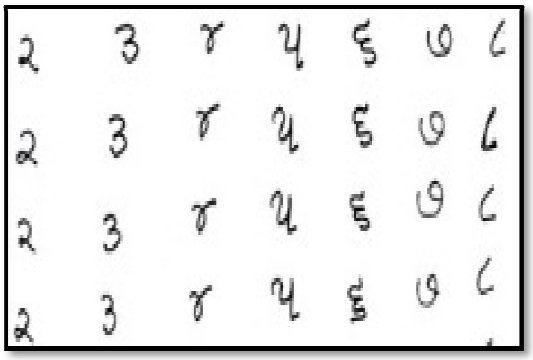


Fig. 3Sample of Handwritten Gujarati Numerals

We have considered here neural network with two hidden layers as for complex problem such as character recognition two hidden layers must be considered. Output neurons are four because for output we have considered four classes as per the BCD coding scheme for individual numeral from 0 to 9 **(i.e. 0- 0000, 1- 0001, 2- 0010, 3- 0011, 4- 0100, 5- 0101, 6- 0110, 7- 0111, 8- 1000, 9- 1001).**

Table 2 MATLAB 14R Implementation (parameters) and Results

METHOD	LENGTH OF FV	INPUT LAYER	HIDDEN LAYER 1	HIDDEN LAYER 2	OUTPUT LAYER	Training Time (in sec)	Recognition Rate (Over all)
DGH	9	9	3	4	4	14	71.8%
CCH	8	8	3	4	4	10	81.9%
QBCCH	32	32	12	4	4	36	98.1%
DGH+CCH (Method A)	17	17	7	4	4	50	94.3%
DGH+QBCCH (Method B)	41	41	15	4	4	60	99.0%

It is clear from above outcomes that Method B using i.e. DGH and QBCCH all together to extract features results in highest recognition rate for Gujarati Numeral as compared to [10-11],[17-18]. We have tested the trained network with 100 new samples of individual numeral to evaluate and estimate the recognition and confusions for individual numeral.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

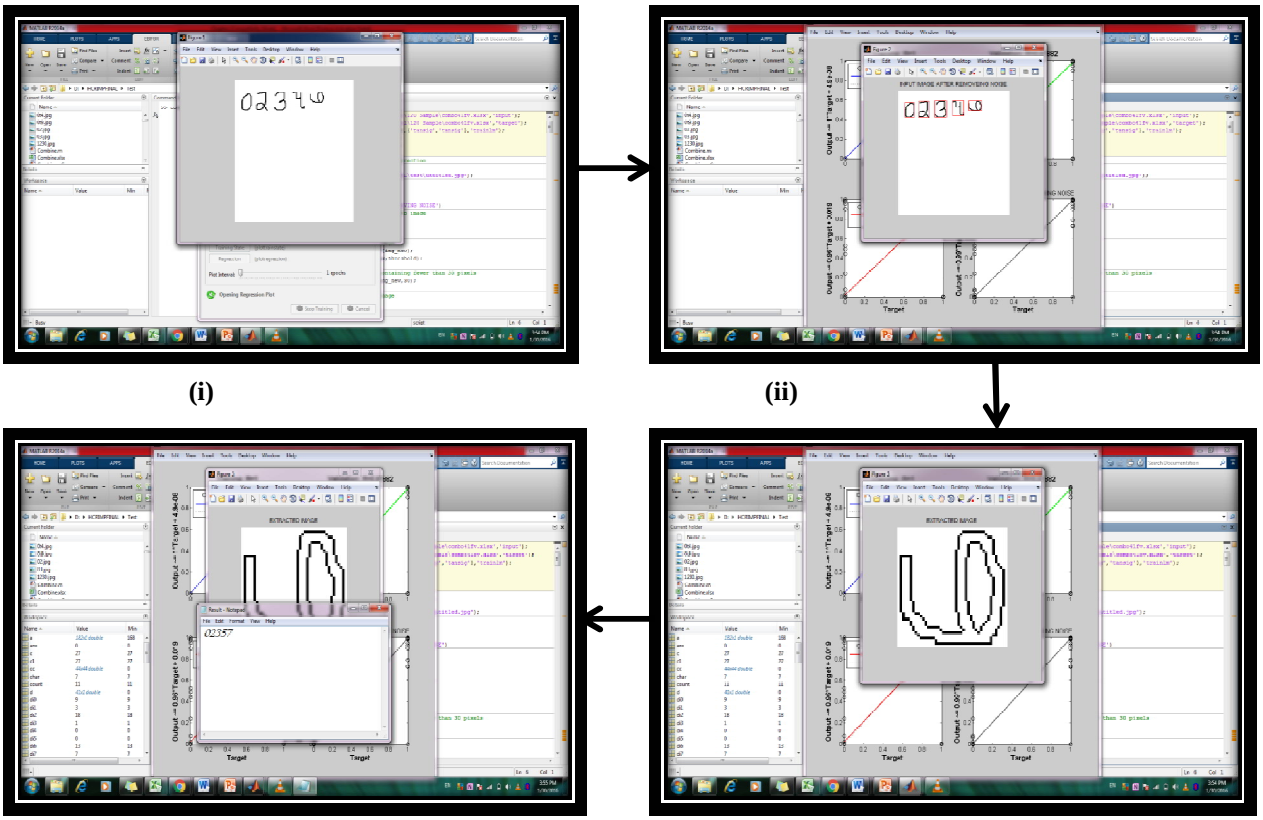
Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

Table.3 Accuracy of Individual Numeral using Method B

Gujarati Numeral	Accuracy (approx.)
0	98%
1	96%
2	96%
3	97%
4	100%
5	100%
6	90%
7	98%
8	97%
9	96%

Table.3 shown below depicts the recognition rate of individual numeral using both directional gradient histogram and quadrant based chain code histogram to extract the features. Training with more number of samples will increase the accuracy of new handwritten numerals. More confusion was found in recognizing Gujarati numeral i.e. 1, 2 and 6. There was no confusion in recognizing 4 and 5.



(i) **(ii)** **(iii)** **(iv)**

Fig.4 Running Scenario using Method B

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

Fig 4(i) Reading and removing noise from the input image file, 4(ii) Segmenting the numeral and normalizing isolated numeral using Method B, 4(iii) Calculating DGH and QBCCH as per Method B, 4(iv) Showing result by writing ASCII in Result.txt with the help of Trained Neural Network

V. CONCLUSION AND FUTURE WORK

It has been concluded with the work that directional gradient feature provides best result only when we consider the feature for each pixel but it results in large feature vector which needs a more complex neural network and computing capabilities. Once the gradient features are normalized to the directional count only the recognition rate falls down. Chain coding is an approach which results in an average recognition rate with the normalized value (feature vector size 8). On dividing the image into quadrants though the size of feature vector gets raised to 32 from 8 but still in less time a good recognition rate can be achieved i.e. approx. 98 %. Recognition rate is comparatively very high in comparison to existing rate i.e. 81.9%. For further enhancing the accuracy the combinational approach was also used to extract feature. Using compound features extracted by directional gradient and chain coding a feature vector of length 17 was created and had resulted in 94.3% of recognition and using directional gradient with quadrant-wise chain coding scheme has raised the recognition to 99% which is superior to other methods. Overall recognition rate using combinational feature vector i.e. Method B is 96.8%. The proposed approach was also tested for the numbers with certain digits.In future work the proposed algorithm can be trained with some other classification algorithm and can be used for some other language.

REFERENCES

[1]Dayashankar Singh, Dr. J.P. Saini and Prof. D.S. Chauhan, "Hindi Character Recognition Using RBF Neural Network and Directional Group Feature Extraction Technique", International Conference on Cognitive Computing and Information Processing, IEEE 2015.
[2]AshutoshAggarwal, Karamjeet Singh and Kamalpreet Singh, "Use of Gradient Technique for extracting features from Handwritten Gurumukhi Characters and Numerals", International Conference of Information and Communication Technologies, Elsevier 2014, pp. 1716-1723.
[3] G Raju, Bindu S Moni and Madhu S. Nair, "A novel handwritten character recognition system using gradient based features and run length count", Indian Academy of Sciences, Vol. 39, Issue 6, December 2014, pp. 1333–1355.
[4]Xianjing Wang and AtulSajjanhar, "Using a Circular grid for Offline Handwritten Character Recognition", 4th International Congress on Image and Signal Processing, 2011, pp. 945-949.
[5]Shen-Wei Lee and Hsien-Chu Wu, "Effective Multiple-features Extraction for Off-line SVM-Based Handwritten Numeral Recognition", International Conference on Information Security and Intelligence Control, IEEE 2012, pp. 194-197
[6] H. Imran Khan, Smitha U.V and Suresh Kumar, "Isolated Kannada Character Recognition using Chain code Features", International Journal of Science and Research, Vol. 2, Issue 8, August 2013.
[7]Qian You, Xichang Wang, Huaying Zhang, Zhen Sun and Jiang Liu, "Recognition Method for Handwritten Digits Based on Improved Chain Code Histogram Feature",3rd International Conference on Multimedia Technology, 2013.
[8] Jomy John, Pramod K. V and KannanBalakrishnan, "Offline Handwritten Malayalam Character Recognition Based on Chain Code Histogram", IEEE, proceedings of ICETECT, 2011.
[9] M.B. Mendapara and M.M. Goswami, "Stroke identification in Gujarati text using directional feature", International Conference on Green Computing Communication and Electrical Engineering, IEEE 2014.
[10]Avani R. Vasant, Sandeep R. Vasant and Dr. G. R. Kulkarni, "Performance Evaluation of Different Image Sizes for Recognizing Offline Handwritten Gujarati Digits using Neural Network Approach", International Conference on Communication Systems and Network Technologies, IEEE 2012, pp. 270-273.
[11]Dr. MamtaBaheti, "Invariant moments approach for Gujarati Numerals", International Journal of Engineering and Applied Sciences, Vol. 2, Issue 2, February 2015, pp. 41-44.
[12]Anitha Mary, M.O. Chacko and P.M. Dhanya , "A Comparative Study of Different Feature Extraction Techniques for Offline Malayalam Character Recognition", International Conference on Computational Intelligence in Data Mining, Springer 2014, Vol. 2, pp.9-18.
[13]Ravindra S. Hegadi and Parshuram M. Kamble, "Recognition of Marathi Handwritten Numerals Using Multi-layer Feed-Forward Neural Network", World Congress on Computing and Communication Technologies, IEEE2014, pp. 21-24.
[14]Aiquan Yuan, Gang Bai, Lijing Jiao and Yajie Liu, "Offline Handwritten English Character Recognition based on Convolutional Neural Network", 10th IAPR International Workshop on Document Analysis Systems, IEEE2012, pp. 125- 129
[15]Dayashankar Singh, MaitreyeeDutta and Sarvpal H. Singh, "Neural Network based Handwritten Hindi Character Recognition System", 2nd Bangalore Annual Compute Conference, Article No. 15, ISBN 978-1-60558-476-8.
[16] Akinori Kawamura, Koji Yura, Tatsuya Hayama, Yutaka Hidai, TadatoshiMinamikawa and Akio Tanaka Shoichi Masuda, "On-line Recognition of Freely Handwritten Japanese Characters Using Directional Feature Densities", Proc. 11th International Conference of Pattern Recognition, Vol. 2, pp. 183-186, 1992
[17] M. J. Baheti and K. V. Kale, "Gujarati Numeral Recognition: Affine Invariant Moments Approach", Special Issue of International Journal of electronics, Communication & Soft Computing Science & Engineering, pp. 140-146, March 2012.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

[18]Hetal R. Thaker and Dr. C. K. Kumbharana, "Analysis of structural features and classification of Gujarati consonants for offline character recognition", International Journal of Scientific and Research Publications, Vol. 4, Issue 8, pp. 1-5, August 2014.

BIOGRAPHY

Nishu Rastogi iscurrently working as Assistant Professor in Department of Computer Applications at Invertis University, Bareilly. She has completed M.E. CSE from National Institute of Technical Teachers Training and Research, Chandigarh affiliated from Panjab University,B.Tech IT from Invertis Institute of Engineering and Technology, Bareilly affiliated from UPTU.

Dr. MaitreyeeDuttaiscurrently working as Professor and Head of Department of Electronics and Communication Engineering at National Institute of Technical Teachers Training and Research, Chandigarh affiliated from Panjab University. PHD in Engineering and Technology from Panjab University, M.Tech in ECE from Panjab University and B.Tech in ECE from Guwahati University.

Mr. Ajay Indianiscurrently working as Associate Professor and Chief Proctor atInvertis University, Bareilly. He is pursuing PHD from GurukulKagadi University, Haridwar, India and M.Tech in CSE from UPTU.